Vocal Learning and Flexibility in the Communication of Common Marmosets (*Callithrix Jacchus)*

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Yvonne Zürcher

aus

St. Gallen SG

Promotionskommission

Prof. Dr. Judith Burkart (Vorsitz)

Prof. Dr. Carel van Schaik

Prof. Dr. Marta Manser

Prof. Dr. Klaus Zuberbühler

Zürich, 2020

Contents

I. Acknowledgement

First and foremost I thank Prof. Judith Burkart for her guidance over the last few years. I am very grateful that I got the opportunity to do this project in your group and under your supervision. I never felt that there was a lack of patience, support and help whenever I was in need. I am grateful for all the discussions and the ample opportunities that let me become the scientist I am today. Thank you also for being a role model in proving that working in Science can be compatible with having a family. I always enjoyed having your kids around!

A big thank you to Prof. Carel van Schaik for all his help, suggestions and support. Both the project and I profited from your endless supply of ideas, questions and discussions! I want to thank Prof. Marta Manser and Prof. Klaus Zuberbühler for being part of my PhD committee, for their time to participate in the committee meetings and helpful comments and suggestions during those. I also want to thank Dr. Claudia Fichtel for reviewing my thesis.

I would like to express my thanks and my gratitude to all the animal care takes that worked at the Primate Station during my time here, but especially to our main care takers: Heinz Galli, Patricia Riveras, and Gisep Bazzell. Nothing we do here would be possible without your never-ending providing of care and without you ensuring the well-being of not only all of our monkeys, but also the humans! I am also thankful to Claudia Zebib, Ruth Haegi and Tony Weingrill who were always answering all the questions concerning the administration and the PhD program, and were always helping to provide solutions to any problems.

I had the pleasure to work with several collaboration partners during the course of my PhD: I owe my thanks to Dr. Augusto Vitale at the Istituto Superiore di Sanità in Rome who allowed me work with his monkeys for two weeks as well as to Prof. Dr. Ana Fidalgo and Prof. Dr. Susanna Sanchez at the Università Autonoma in Madrid that both allowed me to record their colony as well as entrusted us with the four monkeys we translocated to Zürich. I am thankful to the Zoo Zürich for their help with the translocation as well as for providing temporal housing and care for the four monkeys and giving me access to them during that time.

I had several research assistants, to whom I am very grateful: Belinda Kunz, Sonja Foster, Alexandra Kiessling and Rahel Brügger, thank you for your work, handling such an amount of raw data would have been nearly impossible without you!

I was very lucky to write papers with very resourceful co-authors: My thanks go to Dr. Hanna Ruch for initiating the accommodation review, which was a very solid base for my PhD. A huge thank you also goes to Dr. Erik P. Willems, who not only co-authored two of the papers but also was an endless source for statistic support and fundamental for the implementation of many of the analytical ideas I came up with over the years.

My thanks also go to all my colleagues and friends in the PS team. Thank you to Sandro for the proof reading, and to anyone for always being up for a discussion, a quick input or helping with finding answers to questions of any format. A big thank you goes to Rahel, who started her journey at the AIM at the same time as me, and who became a close friend. I would have been lost several times without your support!

But really my thanks go to all my office mates that overlapped with my being at the AIM in space and time: Thank you all for creating an atmosphere that was always supportive and helpful, and thank you for all the laughing!

Finally, I am thankful for all the support from my family and friends that I got during the years working on this project. Knowing that you believe in me and what I do is of enormous value to me! I am especially thankful to my husband, who was at my side during this whole journey and who never stopped reminding me of why I am here in the first place.

Thank you all for making this possible!

II. Thesis Abstract

The evolution of human language has fascinated researchers for a long time. Even though there are no other species with a similar communication system, studying communication and related behaviours in different species seems to be a promising approach as some elements of language are nonetheless present in other species. By comparing traits in closely related species one can investigate which elements were most likely present already in the last common ancestor and which have evolved de novo in the species under study. Studying convergent traits between species that are less closely related but face similar selection pressures allows to investigate patterns that evolved in response to these selection pressures. Both approaches are used to investigate questions about the evolution of human language. Comparing the cognitive capacities of great apes showed that many capacities considered important for human language were most likely already present in the last common ancestor of all great apes. Studying both homologous and convergent traits in additional nonhuman primate species provided evidence that many of the core language elements are present in other species in one form or another as well. Primates use and understand structural rules like syntax or conversational rules during their communication with others and show remarkable vocal flexibility. Even though vocal learning was often considered to be rare in nonhuman primates, looking at vocal learning in more detail reveals that vocal production learning both during infancy as well as in the form of vocal accommodation occurs relatively often. During infancy, vocal learning is best documented in pygmy and common marmosets, where infants are reported to engage in vocal babbling behaviour, and caregivers seem to provide infants with feedback about proper calling behaviour. In adult primates, vocal accommodation seems to be the most common form of vocal learning, indicating that primates can indeed modify their vocalizations due to a social input.

In this thesis, I wanted to investigate the vocal flexibility of common marmosets *(Callithrix jacchus)*. Common marmosets are not only highly vocal and show increasing evidence for vocal learning skills, as cooperative breeders they also share their social system with humans, which makes them a very suitable model system to study convergent patterns in communication. In the first part of the thesis, I investigated whether captive populations of common marmosets do have different vocal dialects. To do so, I compared three different call types between three different captive colonies. I found clear evidence for vocal dialects in each of these three call types. As dialects are consistent with vocal learning but not the only possible explanation, I wanted to establish whether social learning or rather environmental or genetic differences between populations were the most likely explanation for the results I had found. The best method to study such questions are translocation experiments, where animals are moved between different environments or social groups. I found that a new physical environment did not have a long-lasting effect on marmoset vocalizations, but that the introduction into a new social environment did induce vocal accommodation. Four translocated animals became more similar to the new social environment over the time of 16 weeks after introduction in two out of three call types. Vocal accommodation, especially vocal convergence, is often indicating a close social bond or group membership. It can therefore be beneficial for individuals to accommodate to new groups or new partners. If calls though function for individual recognition, strong accommodation could reduce the individual recognisability. Individuals could therefore have different accommodation strategies depending on the call function. To study this question, I formed ten new marmoset breeding pairs and recorded vocal accommodation in three different call

types with different functions. Phee calls are long distance contact calls that are important to transmit individual identity, trill calls are close distance contact calls and food calls are produced by individuals that are willing to share food. I found that animals converged the most with their partners on trill calls, and less so on phee calls. In food calls, I found mainly divergence, indicating that animals are aiming for a larger vocal distance than what they had initially. These results suggest a trade-off between accommodation and individuality, and that animals can use different strategies in different call types to overcome this trade-off. Together, the studies performed during this PhD confirm that calls in marmosets are not genetically fixed and rigid but show a certain flexibility. Common marmosets show social vocal learning in the form of vocal accommodation in different situations and the amount and direction of accommodation depends on the call function. Vocal learning is therefore not absent in primates, as often assumed.

Even though primate vocal communication is profoundly different from human language, we can find many elements that are fundamental to human language also in other primate species. Especially apes show many of the cognitive abilities that are required to manage a complex communication system. Comparing the cognitive capacities, communicative content and communicative attitude between humans, chimpanzees and common marmosets reveals that only in humans, all three of these essential elements are strongly expressed at the same time and therefore likely created the space within which language could evolve.

1. Chapter

-

General Introduction

1.1 Introduction

Language is a human universal found in all human societies around the world. It is unique in its entirety, and according to our current knowledge, no other species has evolved a similarly unlimited and highly cooperative communication system. The question how language evolved in humans has fascinated researchers for a long time, but turns out to be difficult to answer [Christiansen and Kirby, 2003]. One way to study the evolution of a specific trait is a comparative approach. When for example two closely related sister species share a certain trait with each other, but not with a third related species, one can infer that it is most likely that the trait in question evolved since the splitting of the two lineages from the third lineage [Hauser et al., 2002]. The goal with this approach is to distinguish between characteristics that evolved as a new trait since the last common ancestor, traits that were modified over time and those that were preserved since lineages split up. In humans, we cannot study homologues patterns in any sister taxa, as humans are the only recent species of their lineage, but we can look at the next close related lineage, the other great apes. Searching for language homologues in great apes has led to many important insights, in particular that many cognitive elements were quite likely already present in the last common ancestor of humans and apes [Tallerman and Gibson, 2012; Tomasello, 2019]. Still, great apes do not have language, and to study some more language specific questions, a complementary approach to studying homologies is looking at convergence. Convergent traits evolved in different lineages not due to close relatedness, but due to similar selective pressures. These convergences allow us to deduce information about the circumstances under which certain patterns can evolve and be selected for. Again, we cannot directly compare the evolution of language to any other communication system, as there is none similar to language that would allow straightforward comparison. But if we do not look at language as one complex unit, but at the different elements that are essential or characteristic for human language, studying convergent structures in other animal systems can be a promising way to study language evolution [Hauser et al., 2002].

Core elements that specify human language like structural rules (linguistic rules like syntax, or conversational rules), as well as vocal learning and vocal flexibility have been intensely studied in nonhuman animals. Those studies could show that many of these elements exist in one form or another in nonhuman animals as well, often being the result of convergent evolution. I will discuss these different aspects and how they are present in mainly nonhuman primates in the following sections 1.2 and 1.3.

1.2 Structural rules in animal communication

1.2.1 Syntax in animals

Syntax can be defined as a rule to organize elements into combinations where structure is relevant to determine meaning [Berwick et al., 2011; Kako, 1999]. Syntax-like structures and call combinations are found in birds, rock hyraxes and nonhuman primates. In birds, Engesser et al. [2016] found that southern pied babblers combine two signals that are meaningful on their own to a combination with a different, independent meaning. A similar behaviour of combining calls to change the call meaning was found in Japanese tits [Suzuki et al., 2016]. Chestnut-crowned babblers were even found to form calls that are composed of otherwise meaningless elements [Engesser et al., 2019]. Also Rock hyraxes combine song elements and show geographical variation in the call combinations used [Kershenbaum et al., 2012]. In nonhuman primates (hereafter called primates), studies found both evidence for combining two or more

calls into meaningful sequences as well as evidence for modifying calls with a suffix to change call meanings. Diana's- , Campbell's – and putty-nosed monkeys produce call combinations in the context of alarm and contact calling, where combination of calls can lead to different meanings [see Zuberbühler, 2019 for an overview]. Studies about call combination in apes are rare, but there is some evidence for meaningful call sequences as well (Gorillas: [Hedwig et al., 2015], Chimpanzees: [Crockford and Boesch, 2005] Gibbons: [Clarke et al., 2006]).

More species seem to have an understanding of syntax, even though they are not using it themselves. Artificial grammar learning paradigms are often used to test an individual's ability to detect grammar-like structures and dependencies. They show that animals are likely to understand different syntactic structures, but it remains somewhat unclear if they use abstract syntactic rules or some simpler mechanisms to do so. Further, most studies could show that animals have an understanding of finite – state grammars, but fail with more complex grammars [ten Cate and Okanoya, 2012].

To study the language capacity of animals, individuals of different species (usually apes, cetaceans or parrots) were trained to understand and use an arbitrary communication system. Normally, such a system consists either of arbitrary symbols the animals can touch as a mean of communication or of a form of sign language (although sing language has only been used in apes due to morphological restrictions in other species) [Gibson, 2012; Janik, 2012; Pepperberg, 2012]. Studies on these language trained animals are another way to get information about the capacity to understand syntactic rules in these animals. Even though these individuals did not often use syntax to produce sentences, they could clearly understand syntax [reviewed by Gibson, 2012; Kako, 1999; Savage-Rumbaugh et al., 1993]. We can therefore conclude that the capacity to understand syntax is shared with our relatives and did evolve before the use of syntax itself. It seems likely that syntactic understanding evolved in the context of processing social information, as both the processing of syntactic rules as well as social interactions require similar cognitive processes [Seyfarth and Cheney, 2014; Seyfarth et al., 2005].

1.2.2 Conversational rules

Primates do not only understand grammatical rules, vocal exchanges in primates often also follow social, temporal as well as structural rules which are based on socially determined principles [Bouchet et al., 2017]. Pygmy marmosets do not call independently from other individuals, but rather follow a certain order in which the animals are calling one after the other. In a group of three animals, animals were more likely to call only after each other animal had called as well, compared to when one of the animals had not called yet [Snowdon and Cleveland, 1984]. Also in other species communication is regulated by temporal rules such as turn taking, where animals try to avoid overlapping calls with a call partner [Lemasson et al., 2011a; Takahashi et al., 2013]. Squirrel monkeys follow temporal rules for answering a preceding call; calls emitted 0.5s after the first call are considered to be answering calls whereas calls emitted more than 0.5s after the initial call are not [Masataka et al., 1986]. Similar results were found in Japanese macaques with response latencies in a similar range (0.6 – 0.8 s) [Sugiura, 1993]. Common marmosets are well known for their turn-taking behaviour when communicating with each other. When engaging in a call exchange with another individual, common marmosets adapted the timing of their calling to a communication partner in a way that resembled a coupled oscillator, similar to how humans do in conversations (even though on a larger time scale in the marmosets) [Takahashi et al., 2013]. Also Japanese macaques were found to engage in turn taking and adjusted the timing of their response calls

to the calls of their communication partners [Katsu et al., 2018]. Turn taking was also found in non-primates, like meerkats, which avoided call overlaps when calling during group sun bathing, although the call overlap seemed to be regulated by a more simple mechanism like call inhibition when someone is calling already [Demartsev et al., 2018]. Vocal exchange rates and timing of responses can be influenced by both dominance and affiliative bond. Chimpanzees produce pant-grunts exclusively towards higher-ranking individuals [Noë et al., 1980], and chimpanzee females showed a different rate of pant-grunts depending on the presence or absence of both the alpha male and the alpha female [Laporte and Zuberbühler, 2010]. Japanese macaques can vocally converge to a communication partner. When they responded vocally to a coo call, the response call was more similar to the original call than random calls would be [Sugiura, 1989]. Further, Japanese macaques preferably adjusted their calls to the calls of higher-ranking females rather than other social partners [Lemasson et al., 2016]. Other conversational rules can be based on the age of the participants. Campbell's monkeys were more likely to react to the calls of older individuals [Lemasson et al., 2010] and common marmosets responded differently to calls of individuals with different age as well as slightly changed their calling behaviour when getting older [Chen et al., 2009].

It seems that many of these rules are learned by infant and juvenile primates during the first years of their life. As described above, Japanese macaques acoustically match a call when replying to it [Sugiura, 1989]. When presented with playbacks of two calls that were either matched or not matched, juvenile Japanese macaques showed no difference in reaction towards the two playbacks, while adult females showed a clear distinction between the two scenarios. This results indicate that the juveniles did either not yet understand the conversational rule underlying call matching in this species or did not mind the violation of the social norm [Bouchet et al., 2017]. Similar results were found in Campbell's monkeys in a naturalistic context, where young individuals violated conversational rules more often than older individuals, as well as in a playback setting, where younger individuals reacted less on playbacks violating the conversational rules than older individuals did [Lemasson et al., 2011a]. Also in chimpanzees, the typical use of pant-grunts arguably only develops during adolescence, with young infants pant-grunting at any individual they encounter, juveniles almost completely stop pant-grunting and only sub-adult individuals start using pant-grunts as the rest of the adult community [Laporte and Zuberbühler, 2011]. In common marmoset monkeys, not only are behaviours like turn-taking learned during development of infants [Chow et al., 2015], but parents seem to actively provide feedback to young individuals to provide information about the correct behaviour. Young common marmosets interrupted adults more often during turn taking bouts and sometimes responded with a call type inappropriate for the situation. Interestingly, marmoset infants were better at turn-taking with their siblings than with their parents, indicating that the behaviour showed a distinct learning pattern depending on the current communication partner. When infant or juvenile marmosets would interrupt the call of a parent (i.e. ignoring the conversational rule), parents would interrupt the current exchange and did not reply to the infant anymore. Also, when infants were replying with the wrong call types, parents were likely to interrupt this vocalization with a phee call, i.e. the appropriate response. Both of these parental behaviours can be interpreted as feedback provisioning by the adult animals to facilitate the development of the proper calling behaviour in their infants [Chow et al., 2015].

Not just conversational rules can be learned during development, but also several other aspects of vocal communication and vocal productions, which I will discuss in the following section.

1.3 Vocal learning

Vocal learning is a fundamental element of human language. The flexibility of vocal learning is essential to increase the vocal repertoire as well as to adjust the communication system to changes in the environment by creating new expressions [van Schaik, 2016]. There are different forms of vocal learning, which can be divided into contextual learning and production learning. Contextual learning can be further divided into usage– and comprehension learning, where the first describes the process of learning in what situation to use the specific signal, and the latter the understanding of the specific signal in a specific situation [Janik and Slater, 2000] (see Figure 1.1). Vocal production learning (VPL) can also be distinguished in two different categories, namely lexical learning and vocal accommodation. Lexical learning leads to the acquisition of a species specific call repertoire (or lexicon), whereas vocal accommodation leads to a modification of the already existing vocalizations due to a current communication partner (social accommodation) or environmental condition (environmental accommodation) [Ruch, Zürcher and Burkart, 2018] (see Figure 1.1). Both lexical learning and social vocal accommodation depend on the input of a conspecific, which serves as a kind of template [Janik and Slater, 2000; Ruch et al., 2018]. Studies on vocal learning often focus mainly on lexical learning, as learning of novel, probably even artificial sounds is very strong evidence of the vocal learning capacities of a species. But focusing solely on the lexical learning capacities might limit our understanding of the vocal learning capacities of many species [Tyack, 2007]. The more subtle vocal changes seen in vocal accommodation are therefore equally interesting, albeit probably more challenging to detect.

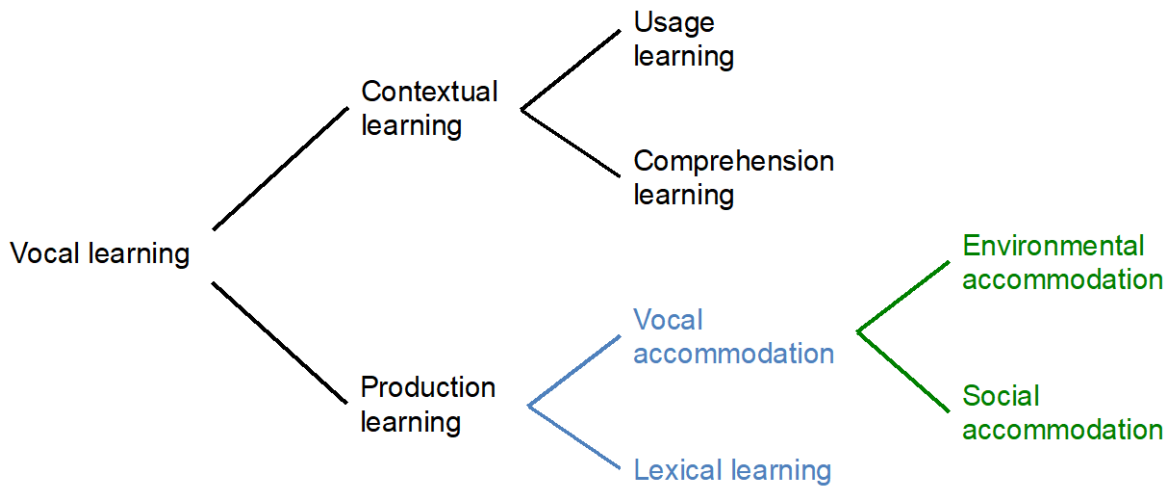

Figure 1.1 An overview over the different forms of vocal learning based on the definition of vocal learning from [Janik and Slater, 2000] and expanded by [Ruch et al., 2018]. Vocal learning can be split up in different categories. The main difference is production learning vs contextual learning. Contextual learning involves the understanding of the call (**comprehension learning**) and when to use it appropriately (**usage learning**), whereas production learning describes the modification of vocalizations to produce a vocalization with specific acoustic properties. Production learning can further be distinguished into **lexical learning** (learning of new call types or words) and **vocal accommodation**, which again can be split up into modifying vocalizations due to a current social partner (**social accommodation**) or a current environmental situation (**environmental accommodation**). Especially lexical learning and social accommodation rely on the vocal input of a conspecific to take place. Figure from Ruch et al. [2018].

VPL in non-human animals is best known and well-studied in birds [Fitch, 2005]. Both songbirds as well as parrots show a large VPL potential and are commonly used to study features of vocal learning. Even though vocal learning in this species evolved independently from humans, mechanisms, neural circuits as well as the genes involved in vocal learning are highly similar between humans and birds as a result of strong convergent evolution [Pfenning et al., 2014]. The famous "language gene" FoxP2 for example shows strong parallels in the expression patterns both in humans and in vocal learning songbirds, indicating that similar mechanisms evolved for a similar function [Scharff and Petri, 2011]. Interestingly, also callitrichids show a comparable expression pattern, indicating that the gene might be involved in a similar function in marmosets as well [Kato et al., 2014].

Evidence for lexical learning in mammals is much rarer. There are some examples of lexical learning in elephants, pinnipeds and bats [Jarvis, 2006; Tyack, 2007]. Therefore, vocal learning abilities are often considered to be absent in non-human primates, especially apes, the sister-taxa of humans, which is quite intriguing considering the immense vocal learning skills of humans. Upon closer inspection, it becomes apparent that lexical learning indeed seems to be absent in nonhuman primates, whereas other evidence for vocal flexibility like babbling infants, vocal accommodation or vocal dialects between different populations are rather common. I will discuss each of these phenomena in more details below.

1.3.1 Vocal learning during infancy in nonhuman primates

Studies on squirrel monkeys showed that infant monkeys reared in vocal isolation did not differ in their vocalizations from normally raised infants, and calls were relatively stable over infancy into adulthood [Hammerschmidt et al., 2001; Lieblich et al., 1980; Winter et al., 1973]. Similar results were found in rhesus macaques that were reared either with or without contact to conspecifics, and where changes in vocalizations were best explained by changes in weight and not by rearing condition [Hammerschmidt et al., 2000]. Cross-fostering infant macaques between species did not lead the infants to develop the calls of the foster species, but the repertoire of their own genetic species [Owren et al., 1992]. Studying gibbon hybrids revealed that songs in this species seem to have a strong genetic foundation, and hybrids tend to show a mixture of the acoustic characteristics of their parental species [e.g. Geissmann, 1984]. The general picture about call development in most nonhuman primates is thus that calls are genetically, rather than socially determined. Humans, in contrast, go through an intensive vocal learning phase in early infancy that starts with extensive babbling behaviour at very early age. This early learning phase is essential for the development of normal speech, and children that – for any reason – do not learn language in this time have trouble learning language at all [Kuhl, 2004].

In contrast to most primates, there is one primate family where babbling and vocal development also seems to be common, namely the callitrichidae. Babbling has been most intensely investigated in pygmy and common marmoset. Similar to humans, the infants of these species begin to produce strings of vocalizations that are clearly different from adult vocalizations shortly after birth. These strings of vocalizations can be rather long and contain both repetitive elements of adult vocalizations as well as infant-typical sounds [Pistorio et al., 2006; Snowdon, 2013]. The amount of babbling is positively correlated with the time adults spend caring for the babbling infant, as well as with the time it takes infants to progress from babbling to a normal adult call repertoire [Elowson et al., 1998a; Snowdon and Elowson, 2001]. Adult common marmoset acting as caregivers also seem to provide infants with feedback

about proper calling behaviour during turn taking (see Chapter 1.2.2) [Chow et al., 2015]. It seems that both babbling and parental feedback are important for marmosets to develop their adult call repertoire. Animals that were separated from their parents at young age and therefore did not get direct vocal feedback retained more infant characteristics in their calls as adults compared to animals with continuous feedback [Gultekin and Hage, 2017; Gultekin and Hage, 2018], and the more feedback infants got, the faster they transitioned from infant-like to adult-like vocalizations [Takahashi et al., 2015; Takahashi et al., 2017]. Great apes have so far not been found to show babbling behaviour, or any other evidence that infants go through a vocal learning phase. A study on agile gibbons suggests though that daughters practice the duets they produce as adults by co-singing with their parents, and that mothers made their songs more stereotypic when daughters co-sang, perhaps to make it easier for those to learn the song [Koda et al., 2013], comparable to motherese in humans [Nelson et al., 1989].

1.3.2 Vocal accommodation

As mentioned before, there is hardly any evidence for lexical learning in primates, which serves as an argument for the absence of vocal learning in primates. A closer analysis of the literature about call modification and vocal flexibility in primates though shows that vocal learning in the form of modifying calls depending on the environment or on current communication partner is rather common in different situations and a variety of species. This phenomenon is called vocal accommodation and is well studied in humans. When we reviewed a large number of studies on vocal accommodation in human and nonhuman primates, we found that mechanism and function of vocal accommodation is highly comparable between the two groups and most probably share an evolutionary history [Ruch et al., 2018].

Vocal accommodation can be classified into four different categories depending on the mechanism and the function of accommodation (see Table 1.1). Individuals can accommodate their vocal output either to their current environment to ensure signal transmission (environmental accommodation), or to a current communication partner (social accommodation). In both cases, accommodation can be regulated by an automatic process (e.g. a perception – production link) or by another (not closer defined) mechanism. Whereas both environmental and social accommodation regulated by an automatic process are common both in humans and primates, examples regulated by other mechanism are much rarer.

Table 1.1 Vocal accommodation can be classified into four different categories depending on the assumed function and mechanism. Vocal accommodation can occur to either optimize signal transmission (A, B) or to express social closeness / distance (C, D), and can be explained either by an automated feedback mechanism (A, C) or not (B, D). We give an example for each form of vocal accommodation both in humans and primates (reference for primate examples in *italics)*. While there are some examples from the primate literature about all four categories, most examples for primates fall into the category C: Social accommodation that can be explained by an automated mechanism like a perception – production link. Table modified after Ruch et al. [2018]

		MECHANISM automatic processes (e.g. perception – production link) likely?	
		yes	**no**
FUNCTION	**Optimizing signal transmission** (environmental accommodation)	**A** Noise-induced increase of amplitude and duration (Junqua, 1996; *Hotchkin & Parks, 2013*)	**B** Speaking/calling louder when communication partner is farther apart (Pelegrín-García *et al.* 2011; *Choi et al., 2015*)
	Expressing social closeness / distance (social accommodation)	**C** Convergence of speech/calls in an interaction (Pardo, 2006; *Candiotti et al., 2012)*	**D** Divergence between members of two social groups (Bourhis & Giles, 1977; *Crockford et al., 2004*

Evidence for environmental accommodation comes for instance from olive baboons, which adapted their grunt vocalizations to environmental conditions by producing longer calls in a closed compared to an open habitat [Ey et al., 2009]. To reduce interference with environmental noise, wild pygmy marmosets produced calls that fell in relatively quiet regions of the frequency spectra of their respective habitats [de la Torre and Snowdon, 2002]. Also captive common marmosets increased their median sound level as well as call duration [Brumm, 2004], and cotton top tamarins modified several frequency parameters of their calls [Hotchkin et al., 2015] in response to increased background noise levels. This so called Lombard effects are widespread in many animals and usually attributed to fully automated mechanisms [Brumm, 2004; reviewed by Hotchkin and Parks, 2013]. This mechanism is supposed to be based on an automatic and bidirectional relationship between the signals an individual receives and produces. The neuronal circuit responsible for integrating vocal production and auditory perception is located in the brain stem in most species and is therefore supposed to be a rather old mechanism working highly automated [Zollinger and Brumm, 2011a; Zollinger and Brumm, 2011b]. Humans though can control the Lombard effect to a certain degree, indicating that also some higher cortical levels are involved in humans [Zollinger and Brumm, 2011b]. Interestingly, there is a similar involvement of prefrontal structures that get activated during the Lombard effect in common marmosets, which could be a hint that also in them vocalizing is to some degree regulated at higher cortical levels [Eliades and Wang, 2012].

Some findings of environmental accommodation are unlikely to be explained by a simple automatic perception-production link, although they are much rarer. They are usually examples from individuals that modify their vocalizations opposite to the changes they perceived. For example, common marmosets increased the amplitude of their phee calls (a long distance contact call) when they were responding to a conspecific from whom they were farther apart (and therefore could hear with a lower amplitude), most likely to prevent turn-taking from breaking down [Choi et al., 2015]. Also Diana monkey females produced calls with more pronounced individual identity when the animals were farther apart from each other and/or the visibility was low [Candiotti et al., 2012].

Social accommodation can take place both over short-term (e.g. accommodating to a current communication partner) or over long-term (e.g. accommodating to a new partner or social group). Japanese macaques adjusted their call frequencies towards a presented playback and converged more towards higher ranking animals [Lemasson et al., 2016; Sugiura, 1989], whereas female Diana monkeys became more similar to a conversation partner during call exchanges [Candiotti et al., 2012]. Also chimpanzee males converged with chorus partners, which was arguably independent of affective states [Mitani and Gros-Louis, 1998]. These studies are all examples of relatively short-term accommodation.

There are also several examples of long-term accommodation in primates. Newly formed pygmy marmoset pairs became more similar towards each other after pairing, and even when calls varied over time, pairs varied their calls in similar ways [Snowdon and Elowson, 1999]. Campbell's monkeys' contact calls were more similar between individuals that shared a higher degree of social affiliation, which was most likely because these animals were more frequently exposed to calls of individuals with whom they were friendly [Lemasson et al., 2011b]. Also, changes in the group composition in these monkeys led to an increased call similarity within groups, probably to increase the signal of group identity [Lemasson and Hausberger, 2004]. In chimpanzees, Watson et al. [2015b] found that after translocating a group of chimpanzees, the animals modified their food grunts and became more similar to the group they were introduced to. These shifts in calls only occurred after the animals had established social bonds with animals of the new group. These results are somewhat controversial [Fischer et al., 2015; Watson et al., 2015a], but can be seen as evidence for vocal convergence or usage learning taking place once social bonds were established. Long-term accommodation is usually described in the situation of group or pair formation, and can mostly be explained by an automatic perception-production link similar to the one responsible for the Lombard effect [reviewed by Ruch et al., 2018].

There are only few examples of social accommodation that could not be explained by an automatic perception production process. Chimpanzee groups produced pant-hoots that are more distinct from the calls of their direct neighbours than from those of random other groups (and therefore seem to diverge from groups in earshot). This indicates that the chimpanzees became more similar to each other while diverging from groups in the vicinity. Interestingly, they were not more different from groups that were further away, indicating that they diverged specifically from neighbours and not just vocal divergence occurring by chance [Crockford et al., 2004]. Wied's black-tufted-ear marmosets modified their call structures after new individuals were introduced into the colony room [Rukstalis et al., 2003]. As in this study the direction of the changes was not investigated, it is hard to say more about mechanisms and function in this example, but it shows that the animals expressed some vocal plasticity.

Vocal accommodation can serve different means. It can help to increase signal transmission, facilitate the recognition of group members and signal pair bond strength. There seems to be a clear link between social accommodation and social closeness in many species [Ruch et al., 2018]. I will examine potential functions of accommodation in more detail in the general discussion. One important benefit seems to be that convergence can facilitate the movement between and integration into new groups [Sewall et al., 2016]. In case the immigrating individual converges to the acoustic properties of the new social group, vocal differences between populations can be stable, and so-called vocal dialects can evolve.

1.3.3 Vocal dialects in animals

Vocal dialects are variations within the species-specific sounds between populations or groups and are often related to vocal learning abilities. Unsurprisingly, dialects are found in many different human languages, as well as in different species of songbirds. Mammals on the other hand show less evidence for different dialects ([Henry et al., 2015; Tyack, 2007], although there are some examples. The probably most prominent representatives are members of the cetaceans like the killer whales [e.g. Filatova et al., 2007; Ford, 1991] and humpback whales [e.g. Noad et al., 2000; Owen et al., 2019; Winn et al., 1981] as well as sperm whales [Rendell and Whitehead, 2003], who are known to produce calls or specific stereotyped call sequences (so called codas) that differ between populations, clans or pods. For nonhuman primates there is rather scarce evidence for acoustic variation between different populations. Vocal differences were found between wild populations of the pygmy marmoset [de la Torre and Snowdon, 2009], Thomas langurs [Wich et al., 2008], chimpanzees [Crockford et al., 2004; Mitani et al., 1992; Mitani et al., 1999] and orangutans [Delgado, 2007], as well as between captive populations of the common marmoset [Zürcher and Burkart, 2017].

As with vocal accommodation, there are several hypotheses about the function of dialect differences. One hypothesis suggests that local dialects allow females to mate with locally adapted males, ensuring that their offspring gain genes that are locally beneficial [Podos and Warren, 2007]. On the other hand, dialect differences might help to avoid inbreeding by indicating the geographic distance to the natal group to a dispersing individual [Jenkins, 1978]. Furthermore, dialects might correlate with social units and seem to function as passwords in birds [Henry et al., 2015; Podos and Warren, 2007], and probably also in mammals [Tyack, 2007]. In this situation, dialects allow individuals to distinguish group members from unfamiliar individuals and might provide means to choose partners that are belonging to the same social unit and / or might be familiar with the local environment due to local knowledge. In human children, similarity in dialects seem to qualify someone as a preferable partner and therefore facilitate cooperation [Haun and Over, 2015].

While dialects can be an indication for vocal learning, there are also alternative explanations for why populations differ in their vocalizations. These are mainly genetic or environmental differences between populations. In the example of the killer whales mentioned above, call variation is highly linked to maternal pods and coda type in sperm whales correlates with mitochondrial haplotype [Tyack, 2007]. It therefore is very difficult to distinguish between genetic causes for dialects and the effect of social vocal learning. Genetic drift can lead to different call structures between populations, as found to be the likely case in elephant seals [Le Boeuf and Petrinovich, 1974]. When elephant seals colonized a new island, the

first animals that arrived on the island seem to have a rather low pulse rate when calling, leading to a characteristic difference between the newly colonized island and the source population. With immigration to the new island continuing, the pulse rate became more similar to the one of the other populations again, indicating that the temporal occurrence of dialects was indeed mainly based on the peculiarities of the founder individuals and not due to environmental differences or social learning [Le Boeuf and Petrinovich, 1974].

Environmental differences can lead to acoustic differences between two groups or populations if they induce environmental accommodation. If animals adapt their calls to the acoustic condition of their current location to increase call transmission, it can lead to population differences in the absence of genetic differences and social vocal learning, given that environmental structures vary enough between different populations. To understand what leads to the observed acoustic differences between populations, it is therefore important to disentangle the contribution of patterns like drift and migration, environmental influences and vocal learning [Hauser et al., 2002]. One highly acknowledged method to do so are translocation experiments, namely moving animals between different populations and environments, to quantify the contributions of the factors mentioned above separately [Laland and Hoppitt, 2003].

1.4 Cooperative aspects of primate communication

As we can see, many elements of human language are present in other species. Human language though requires more than a set of rules and signals. Human communication is highly cooperative on both the side of the sender as well as the receiver [Tomasello, 2010] and is heavily based on social learning, social experience and interaction between individuals. Each of these aspects requires a certain motivational predisposition as well as the cognitive capacity to provide an individual with the possibility and potential to acquire, use and understand language [Burkart et al., 2018]. I will discuss these aspects of communication and how they might have set the stage for language evolution in the discussion section. What I want to stress here is that for language to successfully evolve, high social tolerance and willingness to cooperate was essential. To study elements of language that evolved convergently in other species, it might therefore be highly interesting to look at another species who shares high social tolerance and cooperative motivation with humans. In the realm of the primates, this are mainly the callitrichids, a family of cooperatively living New World primates. The focus of my PhD was therefore on one species of this family, the common marmoset *Callithrix jacchus.*

1.5 Vocal communication in callitrichids

Even though most primates readily use vocal signals to communicate in a variety of situations, the family of Callitrichids is particularly vocal. They are relatively small New World primates native to the forest areas in South America living mainly in forest canopies. Living in dense vegetation with limited visibility might be one of the reasons common marmosets rely more on acoustic signals to communicate than for example gestural or facial communication signals [Snowdon and Ziegler, 2007]. Callitrichids might also have an increased need to communicate due to their lifestyle. Similar to humans, callitrichids are (more or less) monogamous, family-living primates with a cooperative breeding system. Groups of marmosets and tamarins usually consist of a breeding pair, their current infants and several adult, reproductively non-active, related or unrelated helpers. Within group social tolerance is very high, and hierarchies are usually

flat (except for reproductive skew) [Díaz-Muñoz and Bales, 2016]. Within these groups, individuals cooperate in behaviours like vigilance, territory defence and especially infant care by providing different services to the infants like food sharing, grooming or carrying. This increased cooperative behaviours require a high level of group coordination, which supposedly leads to an increased need for communication [Burkart et al., 2018; Snowdon and Ziegler, 2007]. Indeed, marmosets and tamarins are found to have not only remarkably large vocal repertoires [Burkart et al., 2018] which also includes a variety of call combinations, but also a very high vocal output. They show an astonishingly high vocal flexibility and increasing evidence for the potential of vocal learning, both during infancy [Choi et al., 2015; Chow et al., 2015; Pistorio et al., 2006; Snowdon, 2013; Takahashi et al., 2016; Takahashi et al., 2015; Takahashi et al., 2017] as well as when adult [Elowson and Snowdon, 1994; Roy et al., 2011; Rukstalis et al., 2003; Snowdon and Elowson, 1999; Zhao et al., 2019]. From the examples discussed in Section 1.3.1, vocal flexibility during development is almost exclusively reported for callitrichid monkeys. Also the reports for vocal flexibility in adults (Section 1.3.2 and 1.3.3) contain a high number of studies on marmosets or tamarins.

Their predisposition for vocal learning, their high vocal output as well as their social system that is comparable to the human social system makes callitrichids a highly valuable study system to investigate questions about human evolution. Even though they are less closely related to humans as for example chimpanzees and bonobos, the social constraints they faced during evolution might have been more similar to those in humans in certain aspects. Common marmoset are increasingly used as a model system to study cognition and communication both on a behavioural or neuronal level [Burkart and Finkenwirth, 2015; Marx, 2016]. Their neural mechanism controlling vocalizations is similar to the one in humans, which makes them a suitable system to study the neural basis of human language [Kato et al., 2014]. Similar to humans, they start vocalizing at a very early age and their vocalizations mature during the development into the adult vocal repertoire. As adults, they remain vocally flexible to a certain degree and show vocal learning potential in the form of vocal accommodation, as shown throughout this thesis. Due to their easy handling and breeding, they are relatively abundant and housed at different research institutes. This makes it possible to compare captive populations between different facilities as well as to exchange animals between different facilities.

1.6 Goals and questions

The main goal of my PhD thesis was to investigate the vocal learning potential of common marmosets. Given the information found in the literature, I assumed that common marmosets are capable of a certain amount of vocal learning, especially vocal accommodation. This led us to the expectation that different populations of common marmosets should vary in their vocalizations. My first step was therefore to compare vocalizations between different populations. I recorded common marmoset in three captive colonies and compared three call types between the three locations, finding that they indeed had population level differences in call structure (Chapter 2). As the occurrence of population differences is consistent with social vocal learning but does not prove it, I went on to investigate how different factors like environmental or genetic differences between populations could have contributed to the vocal dialects I found. One of the best methods to test potential causes of population differences are translocation experiments [Laland and Hoppitt, 2003]. I therefore translocated common marmosets both between different housing conditions (different physical environments) as well as between different

social groups. Over the whole translocation process, I recorded the vocalizations of the animals and analysed which change in external condition (new environment, new social background) would lead to changes in the vocal structure of the animals. I found that changes in the social background, but not in the physical environment led to long-lasting changes in vocal structure, indicating that common marmosets show vocal learning in the form of social accommodation (Chapter 3). Vocal accommodation can signal group membership and a social bond, but it can also reduce individuality in calls. This could lead to a trade-off between the desired function of accommodation and the function of the call per se, especially when the latter serves for individual recognition. Therefore, I would expect a different pattern of social accommodation depending on the call function, in particular with regards to the importance of signalling identity within a given call. To investigate this question, I followed the vocal development of 20 marmosets after forming new breeding pairs. In this set up, I could focus in more detail on the questions of who was accommodating to whom, and how accommodation might change due to different call types and their usage. I analysed three different call types with different functions: phee calls, a long distance contact call, trill calls, a short distance social call as well as food calls, which are usually produced when animals are indicating their willingness to share food. I hypothesised that identity is more important for phee calls and food calls and less important for in trill calls, which would allow more convergence in the latter. Indeed, I found strongest evidence for convergence in trill calls, accompanied by a decrease in individual identity encoded in the calls. I found less convergence in long distance contact calls and individual identity remained stable at a high level in this call type. I found mainly divergence in food calls, indicating that animals aimed for a larger vocal distance from their partner than what was initially present (Chapter 4). At the end I give a short overview over potential communication content, cognitive capacity and communication attitudes in humans and both chimpanzees and common marmosets to discuss potential reasons why only in humans language evolved (Section 5.4).

2. Chapter

-

Evidence for dialects in three captive populations of common marmosets (*Callithrix jacchus*)

Yvonne Zürcher[1] & Judith M. Burkart[1]

[1]Department of Anthropology, University of Zurich, 8057 Zürich, Switzerland

Yvonne Zürcher
yvonne.zuercher@uzh.ch

Published in: International Journal of Primatology, 38(4), 780-793.

2.1 Abstract

The vocal repertoires of nonhuman primates have long been thought to be invariable across populations and not to result from vocal learning. However, increasing evidence suggests that learning does influence vocal production in nonhuman primates, and that several species modify the structure of their calls in response to social or environmental influences. Vocal usage learning refers to the process where an individual learns in which circumstances to produce a certain call type, whereas vocal production learning refers to the process where signals get modified as the result of individual experiences. Common marmosets (*Callithrix jacchus*) show socially mediated vocal plasticity as adults and during vocal development. This propensity to engage in simple forms of vocal production learning (accommodation) should produce population-level differences in call structure. To test this prediction, we compared the vocalizations of three captive populations of common marmosets. We analysed the acoustic structure of 1337 phee calls, 461 trills and 3611 food calls and compared them with a permutated discriminant function analysis. We found that all call types differed significantly between the three populations, and 76-98% of the calls were correctly classified. As physical differences in body mass and environmental differences between colonies could not explain the call differences, we conclude that vocal accommodation is the most likely explanation for the differences in call structure. This will allow us to further investigate the role and importance of vocal learning in a species increasingly used to study vocal learning and language evolution.

Key words: *vocal accommodation, dialect, Callithrix jacchus, vocal learning*

2.2 Introduction

Most mammals, in contrast to humans, generally do not rely on vocal learning to acquire their species specific vocal repertoire. Vocal learning can be divided into vocal usage learning and vocal production learning. Vocal usage learning refers to the process where an individual learns in which circumstances to produce a certain call type, whereas vocal production learning refers to the process where signals get modified as the result of individual experience [Janik and Slater, 2000].

Vocal repertoires in nonhuman primates have traditionally been thought to be genetically fixed rather than to result from vocal production learning [Egnor and Hauser, 2004], and therefore to show little variation within and between populations. For instance, immature squirrel monkeys (*Saimiri sciureus*) and rhesus macaques (*Macaca mulatta*) can acquire species-specific adult repertoires without auditory input from conspecifics [Hammerschmidt et al., 2001; Hammerschmidt et al., 2000; Lieblich et al., 1980; Winter et al., 1973]. Gibbon vocalizations also seem to be under strong genetic control, as studies of gibbon hybrids suggest limited flexibility [Bockelman and Schilling, 1984] [but see Koda et al., 2013]. However, vocal production learning can also occur in the subtler form of accommodation, in which call structure changes due to environmental or social influences. Whereas evidence for acquiring new call types is rare in nonhuman primates, accommodation is more widespread [Candiotti et al., 2012; de la Torre and Snowdon, 2002; de la Torre and Snowdon, 2009; Egnor and Hauser, 2004; Hotchkin and Parks, 2013; Lemasson and Hausberger, 2004; Ruch et al., 2018; Rukstalis et al., 2003].

Callitrichids (i.e., marmosets and tamarins) are highly vocal and several studies suggest remarkable vocal plasticity [Rukstalis et al., 2003; Snowdon, 2013]. Adult marmosets adjust their call structures to new mates or change calling patterns due to auditory contact with new neighbours [Elowson and Snowdon, 1994; Rukstalis et al., 2003; Snowdon and Elowson, 1999]. Moreover, in contrast to most other nonhuman primates, where vocal plasticity is mainly found in adults [Egnor and Hauser, 2004], vocal production learning has also been reported during the development of immatures in callitrichids. Pygmy marmoset (*Cebuella pygmaea*) and common marmoset (*Callithrix jacchus*) infants show a babbling phase, which accelerates and defines the acquisition of the adult calls [Pistorio et al., 2006; Snowdon, 2001; Snowdon and Elowson, 2001]. Common marmoset infants rely on adult feedback to learn conversational rules like turn taking [Chow et al., 2015; Takahashi et al., 2016] and adult-like call structure [Snowdon, 2001; Takahashi et al., 2015]. Accordingly, a deaf marmoset infant [Roupe et al., 2003] and infants that were deprived of parental feedback by separation [Gultekin and Hage, 2017] had problems developing proper adult calls and continued using infant forms until after adolescence.

We examined population differences in captive common marmosets. To do so, we examined a broad array of acoustic parameters of three call types in common marmosets to test the hypothesis that the propensity to engage in vocal accommodation to social partners should lead to population differences in call structure, or dialects. We focused on the three most common social call types, namely phee calls, trill calls and food calls [Bezerra and Souto, 2008; Vitale et al., 2003]. Phee calls are long distance contact calls usually made when animals are separated from each other, and used in turn-taking sequences between animals [Takahashi et al., 2013]. Trill calls are close distance social contact calls which are often emitted during foraging in the wild and food calls are emitted when animals find and communicate about preferable food [Bezerra and Souto, 2008; Vitale et al., 2003]. These calls are used in similar contexts as

the J and trill calls of pygmy marmosets which were found to differ between populations in the study from de la Torre and Snowdon [2009]. Alarm and mobbing calls are also emitted frequently but are expected to be less flexible, as it is relevant to ensure immediate understanding [Schibler and Manser, 2007]. Based on marmosets' propensity to engage in vocal accommodation to social partners and evidence for population differences in five acoustic parameters of two call types in wild pygmy marmosets [de la Torre and Snowdon, 2009], we hypothesised that population differences in call structure, i.e. dialects would develop in separated populations. We controlled for possible effects of breeding status and body mass. Based on our hypothesis, we predicted that different populations would show differences in the structure of phee calls, trill calls, and food calls.

2.3 Methods

We recorded three call types (phee calls, trill calls, and food-associated calls), from 55 adult common marmosets from three captive populations at the Istituto Superiore di Sanità in Rome (IT), the Universidad Autónoma de Madrid (ES) and the Department of Anthropology, University Zürich (CH) (Table 2.1). These populations have never exchanged individuals.

Table 2.1: Overview of the composition of 15 groups of common marmosets and sample composition for this study (total N=55). The number before the brackets indicates the number of individuals per group belonging to each status/sex class living in that group, the figures in brackets are the number of phee calls, trill calls and food calls included in the analyses. We did not include vocalizations from juveniles and infants. We collected data in February 2014 in the Istituto Superiore di Sanità in Rome, Italy (14 individuals), in 2014 and 2015 at the University of Zürich, Switzerland (25 individuals) and in October-November 2015 at the Universidad Autónoma de Madrid, Spain (16 individuals).

Population	Group	Breeder female	Breeder male	Helper female	Helper male	Juveniles / infants
Madrid	C-Females	-	-	3*(64,32,73)*	-	-
	Woody	1*(11,16,93)*	1*(9,6,98)*	3*(167,40,155)*	1*(40,36,115)*	1
	Wonka	-	1*(12,29,-)*	-	7*(202,99,444)*	-
Rome	c	1*(-,-,180)*	-	-	1*(-,-,59)*	-
	e	1*(-,-,282)*	1*(7,-,25)*	1*(-,-,-)*	-	-
	f	1*(-,-,99)*	-	1*(-,-,33)*	-	2
	g	1*(-,-,11)*	-	3*(64,-,185)*	-	-
	v	1*(-,-,-)*	1*(-,-,111)*	-	2*(88,-,185)*	2
Zürich	Jojoba	1*(11,8,111)*	1*(13,32,38)*	3*(158,16,254)*	2*(60,15,180)*	-
	Kyros	-	-	-	1*(33,26,63)*	-
	Lancia	1*(44,14,91)*	1*(33,-,119)*	3*(81,30,154)*	1*(41,-,141)*	-
	Mia	1*(10,28,76)*	1*(12,-,92)*	-	-	-
	Mina	1*(-,-,-)*	1*(27,9,-)*	1*(-,-,-)*	3*(48,13,115)*	-
	Nina	1*(-,-,-)*	1*(17,12,-)*	2*(16,-,29)*	3*(51,-,-)*	-

Animals were housed in pairs or in groups, except for one animal in Zürich (Kyros) who was housed alone for a short time for husbandry reasons. All animals were kept in accordance with the respective national legislation. For details of housing conditions in Rome see Vitale et al. (2003), and Finkenwirth et al. (2015)

for Zürich (CH). Animals in Madrid (ES) were kept in outdoor enclosures with unlimited access to a wooden, sheltered compartment which was heated (comparable to an indoor enclosure). The cages were separated by several meters and did not allow visual contact. Cages were equipped with wooden branches and enriched with ropes, hammocks and other devices. The diet was comparable between the institutes and included morning porridge and a mixture of fruits and vegetables around midday, enriched with various types of animal protein.

2.3.1 Recording procedures

In Rome, we recorded the animals mainly individually. The recording cage was approximately 1.0 x 1.5 x 1.5 m with a fiberplex front and three wire mesh sides. The test cage was in the same room as group f, but visually separated. All animals were familiar with the test cage, as they had regular access to it as an extension of their home cages. We positioned the microphone in front of the one wire mesh wall that did not face the back wall and was not adjacent to the next cage. We provided food in front of the microphone inside the cage. We made recordings in February 2014 over a period of two weeks, and recorded each individual once or twice a day.

In Madrid, we recorded the animals mainly in pairs and only separated them to elicit phee calls. We separated pairs of animals from the group in a wooden shelter, which had an approximate size of 1.5 x 1.5 x 2.0 m and directed the microphone directly at the focal individual. To identify the caller, we marked the calls directly on the recording using a tagging function. Recordings lasted around 20 min and we changed focal animal after the first 10 min to record two animals per session. We recorded during several days in a two week period in October/November 2014.

In Zürich, we recorded the individuals either in pairs in their home cage in the same way as in Madrid, or alone in a separate familiar testing cage as in Rome. The animals were habituated to the separate testing cage before the recording sessions started. We identified and marked the caller directly on the recording as in Madrid.

We distributed the recording sessions between 9 am to 5 pm. To obtain phee calls, we separated animals (individually or in dyads) from the group with auditory but not visual contact. Trill calls were produced and recorded without intervention. To obtain food calls we provided the animals with a mixture of highly preferred food items. These food items varied between colonies according to local preferences, which varied considerably between colonies. By providing locally highly preferred food, we ensured that the animals had a similar level of excitement. We deduced food preferences either from observations during feeding or from the local animal keepers' knowledge. We provided the animals with as much food as they would take or until we had recorded sufficient vocalizations (approximately 300 calls per session). We used an Avisoft UltraSoundGate 116H and a Condenser Microphone CM16/CMPA for all recordings.

2.3.2 Ethical Note

We performed our study in accordance with the respective national laws. The study in Zurich was approved by the Kantonales Veterinäramt, license number 183/13 24826. To prevent stress to the animals, we interrupted recording sessions when animals showed any signs of distress (e.g. pilo-erection of the tail, attempts to leave the cage).

2.3.3 Processing of the recordings

We visually evaluated and categorized each call into 8 categories. We only included calls that could be clearly assigned to phee calls, trills, or food calls in the analysis. We categorized calls as *unknown* if we could not classify them unequivocally. We combined the calls from all recording situations but excluded calls if they were too soft or too short to be measured by Praat (see below), or if some other noise (other call, moving animal, background) interfered with it.

2.3.4 Statistics

We measured 16-18 acoustic parameters per call using Praat [Boersma and Weenink, 2009; script by Reby and McComb, 2003, adapted by E.F. Briefer] (Table 2.2). We inspected each Praat measurement manually to exclude measuring errors and then performed an unrotated Principal Component Analysis (PCA) for each call type. To compare the call structure between the colonies, we included each PCA factor with an Eigenvalue >1 in a permutated Discriminant Function Analysis (pDFA), with *individual* as control factor and *sex*status* as restriction factors [Mundry and Sommer, 2007]. To identify factors which accounted for differences in overall call structure, we used Linear mixed effect models (LMEs) as post hoc tests with the PCA factors as response, *colony* as fixed factor and *identity* nested within *family* as random factor. We considered the model with the lowest AIC as the best model; differences in AIC bigger than 7 were considered meaningful.

Table 2.2: Description of the acoustic parameters we used to quantify the structure of three call types in common marmosets, based on Briefer and McElligot [2011].

Factor	Description
F0Start	Frequency value of fundamental frequency at the start of the call
F0End	Frequency value of fundamental frequency at the end of the call
F0Mean	Mean fundamental frequency across the call
F0Min	Minimal of the fundamental frequency across call
F0Max	Maximum of the fundamental frequency across call
TimeF0Max	Percentage of total call duration when fundamental frequency was maximum
F0AbsSlope	Absolute slope of the fundamental frequency within the call
F0 Var	Mean variation of the fundamental frequency per second
Fm Rate	Number of complete cycles of fundamental frequency modulation per second (only for trill calls)
FM extent	Mean peak-to-peak variation of each fundamental frequency modulation (only for trill calls)
Q25%	Frequency values at the first quartiles of energy
Q50%	Frequency values at the second quartiles of energy
Q75%	Frequency values at the third quartiles of energy
Fpeak	Highest frequency of the whole spectrum
%Time of max intensity	Percentage of time Fpeak is reached
Jitter	Mean absolute difference between frequencies of consecutive F0 periods, divided by the mean frequency of F0
Shimmer	Mean absolute difference between the amplitudes of consecutive F0 periods divided by the mean amplitude of F0
Sound duration	Duration from beginning to end of the call in seconds

Since animals from different colonies differed in body mass, we tested whether body mass or colony better predicted differences in call structure. We calculated LMEs with the PCA factors as response variables, *body mass* and *colony* as fixed factors, and *breeding/sex-status* and *identity nested in family* as a random factor. We chose the best models according to the AIC.

Our sample size was unbalanced because we had different numbers of calls for classes of animals from different colonies. The permutation step of the pDFA controls for this by producing random, balanced subsamples for each randomization. We tested for model fit and normality, linearity and heteroscedasticity of residuals and found that the data met the requirements.

2.4 Results

We analysed 3611 food calls, 1319 phee calls and 461 trill calls. For each call type, we extracted 5 PCA factors with an Eigenvalue larger than 1 from the parameters measured in Praat (Table 2.3).

Table 2.3: Factor loadings of acoustic parameters on five PCA factors (F1-F5) with an Eigenvalue >1 extracted for three call types in common marmosets. Factor loadings larger than + 0.5 or smaller than – 0.5 are marked in bold, indicating that the parameter made a considerable contribution to the factor. For description and definition of the parameters see Table 2. We collected data in February 2014 in the Istituto Superiore di Sanità in Rome (Italy), in 2014 and 2015 at the University of Zürich (Switzerland) and in October/November 2015 at the Universidad Autónoma de Madrid (Spain).

	Call type														
	Phee					**Trill**					**Food**				
Parameter	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
sound_duration	0.29	**-0.76**	0.01	0.11	0.14	-0.16	**0.54**	-0.12	**-0.59**	-0.18	0.16	0.42	-0.20	0.24	-0.31
MeanF0	**0.97**	0.06	0.07	0.12	-0.01	**0.94**	0.26	-0.15	-0.01	0.02	**0.98**	-0.02	-0.13	-0.13	-0.01
F0start	**0.81**	0.19	-0.39	0.05	-0.15	**0.93**	-0.11	-0.14	0.13	-0.07	**0.97**	-0.03	0.05	-0.11	-0.11
F0end	**0.83**	0.10	0.22	0.32	0.13	**0.82**	0.39	-0.21	-0.15	0.11	**0.91**	0.04	-0.35	-0.13	0.05
MaxF0	**0.92**	0.06	0.22	0.21	0.05	**0.88**	0.42	-0.08	-0.05	0.00	**0.98**	-0.03	0.03	-0.11	-0.07
TimeMaxF0	-0.14	-0.26	**0.50**	0.28	**0.51**	-0.35	0.36	-0.18	-0.43	0.31	-0.06	-0.35	-0.39	0.08	**0.61**
MinF0	**0.86**	0.13	-0.35	0.09	-0.07	**0.94**	-0.10	-0.25	0.05	0.03	**0.91**	0.04	-0.34	-0.13	0.02
F0absslope	0.00	**0.73**	**0.61**	-0.09	-0.09	0.03	0.38	**0.57**	**0.58**	0.11	0.43	-0.30	**0.69**	-0.13	0.05
F0Var	-0.01	**0.70**	**0.68**	-0.02	-0.08	0.05	**0.56**	**0.67**	0.30	-0.02	0.30	-0.31	**0.77**	-0.24	-0.08
FmRate	-	-	-	-	-	-0.04	**-0.72**	-0.02	0.04	0.02	-	-	-	-	-
FMextend	-	-	-	-	-	-0.05	**0.83**	0.34	0.02	0.00	-	-	-	-	-
Q25	**0.83**	0.04	-0.05	-0.13	0.11	0.48	0.05	-0.37	0.35	-0.25	**0.88**	-0.03	-0.06	0.21	0.03
Q50	**0.80**	0.09	0.08	-0.39	0.17	-0.39	0.18	-0.44	**0.52**	-0.30	**0.56**	-0.08	0.20	**0.68**	0.16
Q75	0.20	0.05	-0.11	**-0.84**	0.27	-0.25	0.08	-0.40	**0.50**	0.19	0.14	-0.05	0.21	**0.88**	0.09
Fpeak	0.08	-0.41	0.02	0.17	0.04	-0.04	0.24	0.01	-0.29	-0.39	0.07	0.26	0.01	-0.03	-0.25
Timeofmaxintensity	-0.08	-0.07	0.07	-0.06	**0.71**	0.08	0.07	-0.15	0.02	**0.79**	0.01	**-0.59**	-0.05	-0.22	0.47
jitter	-0.24	**0.68**	**-0.52**	0.22	0.31	**0.55**	-0.47	**0.52**	-0.08	-0.03	0.08	**0.84**	0.24	-0.18	0.39
shimmer	-0.27	**0.69**	-0.48	0.22	0.30	0.47	**-0.50**	**0.55**	-0.16	-0.01	0.16	**0.82**	0.26	-0.06	0.43

All call types differed significantly between populations (Figure 2.1).

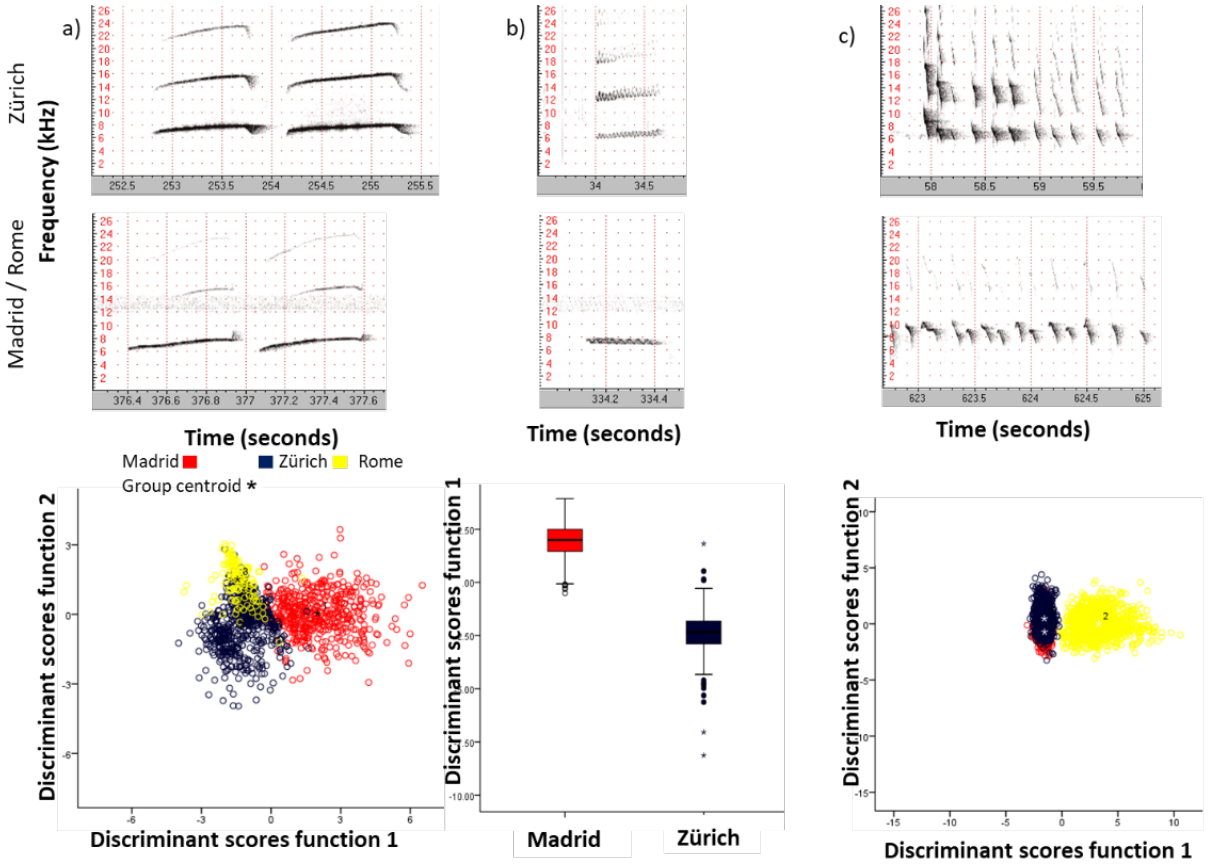

Figure 2.1: **Spectrograms of three call types in common marmosets** (first and second row) **and results of Discriminant Function Analysis to test for population differences** (third row). The first row of the spectrograms shows calls from helper males in Zürich, the second row from a breeding male in Madrid (phee and trill) or from a helper male in Rome (food call). We selected calls close to the group centroid in the discriminant function plots to be most representative. The pDFA results show the first and second discriminant function for a & c and the first discriminant function in b. Populations are: red = Madrid, yellow = Rome, dark blue = Zurich. a) phee calls, p=0.001, b) trill calls, p=0.001 (only Madrid and Zürich due to a too small sample from Rome); c) food calls, p=0.001. We collected data in February 2014 in the Istituto Superiore di Sanità in Rome (Italy), in 2014 and 2015 at the University of Zürich (Switzerland) and in October/November 2015 at the Universidad Autónoma de Madrid (Spain). Boxplots show the median, interquartile range, outliers as circles and extremes as asterisk.

The proportion of calls of each type assigned to the correct colony was significantly above the expected level calculated by the pDFA (Table 2.4). The correct classifications ranged 76-98%, which corresponded to 16-40 % more correct classifications than expected by chance.

Table 2.4: Results of a permutated Discriminant Function analysis (pDFA) testing for population differentiation in common marmoset vocalizations. We obtained expected values by calculating the mean of correctly classified calls in the randomized datasets created in the pDFA. We collected data in February 2014 in the Istituto Superiore di Sanità in Rome (IT), in 2014 and 2015 at the University of Zürich (Switzerland) and in October/November 2015 at the **Universidad Autónoma de Madrid (Spain).**

Call type	Comparison	n	Correct (%)	Expected (%)	p-value	Crossvalidated correct (%)	Crossvalidated Expected (%)	p-value
Phee	Madrid – Rom – Zürich	1304	85.1	52.2	0.001	81.6	40.0	0.001
	Madrid – Rom	655	95.6	66.6	0.001	95.4	58.4	0.001
	Zürich – Rom	806	83.4	66.7	0.005	78.9	54.4	0.003
	Zürich – Madrid	1153	95.3	62.0	0.001	95.6	56.8	0.001
Trill	Zürich – Madrid	461	98.0	62.5	0.001	96.7	59.4	0.001
Food	Madrid – Rom – Zürich	3598	81.6	46.8	0.001	79.5	43.1	0.001
	Madrid – Rom	2137	97.9	64.4	0.001	97.2	63.6	0.001
	Zürich – Rom	2622	98.3	62.9	0.001	98.1	58.6	0.001
	Zürich – Madrid	2437	76.0	59.1	0.001	73.1	55.9	0.001

Post hoc linear models showed that the colonies differed significantly in four (food calls) or five (phee calls and trill calls) of five PCA factors (Table 2.5).

Table 2.5: Results of post-hoc analyses testing which Principle Components contributed most to differences in three call types between three populations of common marmosets. Values p<0.05 are marked in bold. (df = degrees of freedom). We collected data in February 2014 in the Istituto Superiore di Sanità in Rome (Italy), in 2014 and 2015 at the University of Zürich (Switzerland) and in October/November 2015 at the Universidad Autónoma de Madrid (Spain).

	Phee			**Trill**			**Food**		
	df	f	p	df	f	p	df	f	p
Factor 1	41.786	11.372	**0.000**	38.266	3.149	0.54	41.796	11.079	**0.000**
Factor 2	40.005	16.837	**0.000**	34.641	234.097	**0.000**	39.872	18.703	**0.000**
Factor 3	42.37	7.452	**0.002**	38.077	4.494	**0.018**	42.121	5.020	**0.011**
Factor 4	42.051	16.303	**0.000**	37.441	14.619	**0.000**	42.034	14.234	**0.000**
Factor 5	41.504	6.561	**0.003**	38.236	16.569	**0.000**	41.435	8.25	**0.001**

The monkeys from Zürich were significantly heavier than the monkeys from Madrid (LME, fixed factor: colony, random factor: sex*status, F = 26.6, df = 30.21, p < 0.001; no body mass data were available for Rome). However, none of the best models explaining variation in each PCA factor contained *body mass*. The best models all included *colony* as a single fixed factor (dAIC < 7, Table 2.6).

Table 2.6: Result of linear mixed effect models testing for the influence of *colony, body mass, colony&body mass,* and *colony&bodymass* on five PCA factors in three different call types in common marmosets. P values < 0.5 are bold, and letters in brackets indicate which factor of the model they belong to ([c] = colony, [m] = mass, [c*m] = interaction between colony and mass). ΔAIC values are calculated as the distance of each model to the respective best model. For description and definition of the parameters see Table 2. We collected data in 2014 and 2015 at the University of Zürich (Switzerland) and in October/November 2015 at the Universidad Autónoma de Madrid (Spain).

		Model							
		[colony]		[mass]		[colony], [mass]		[colony], [mass], [colony]*[mass]	
Call type	factor	ΔAIC	p-value	ΔAIC	p-value	ΔAIC	p-value	ΔAIC	p-value
Phee	fac1	**0**	0.143	10.692	0.326	9.88	0.78 [m] 0.213 [c]	18.098	0.705 [m] 0.904 [c] 0.772 [c*m]
	fac2	**0**	**<0.001**	22.52	**0.002**	9.924	0.883 [m] **<0.001 [c]**	18.187	0.506 [m] 0.85 [c] 0.895 [c*m]
	fac3	**0**	**<0.001**	13.218	**0.001**	8.043	0.137 [m] **0.016 [c]**	15.909	0.086 [m] 0.218 [c] 0.368 [c*m]
	fac4	**0**	**<0.001**	11.133	**0.001**	8.171	0.137 [m] 0.062 [c]	15.032	0.329 [m] 0.524 [c] 0.212 [c*m]
	fac5	**0**	**0.04**	13.517	0.702	8.366	0.124 [m] **0.011 [c]**	14.509	0.525 [m] **0.041 [c]** 0.086 [c*m]
Trill	fac1	**0**	**<0.001**	23.944	**0.009**	9.218	0.401 [m] **<0.001 [c]**	10.46	0.935 [m] **0.001 [c],** **0.008 [c*m]**
	fac2	**0**	**<0.001**	14.364	**<0.001**	6.548	0.053 [m] **0.004 [c]**	12.123	0.143 [m] 0.17 [c] 0.068 [c*m]
	fac3	**0**	**<0.001**	20.032	**0.003**	10.18	0.984 [m] **0.002 [c]**	18.648	0.892 [m] 0.96 [c] 0.649 [c*m]
	fac4	**0**	**0.043**	12.696	0.39	9.323	0.366 [m] **0.046 [c]**	15.365	0.174 [m] 0.188 [c] 0.105 [c*m]
	fac5	**0**	0.896	9.391	0.907	10.052	0.753 [m] 0.753 [c]	18.58	0.837 [m] 0.708 [c] 0.738 [c*m]
	fac1	**0**	0.862	9.288	0.993	9.848	0.81 [m] 0.769 [c]	18.308	0.808 [m] 0.931 [c] 0.898 [m*c]

Food	fac2	**0**	0.106	10.488	0.231	11.227	0.808 [m] 0.272 [c]	20.935	0.822 [m] 0.797 [c] 0.703 [c*m]
	fac3	**0**	**0.003**	12.471	**0.01**	9.985	0.953 [m] 0.082 [c]	17.382	0.996 [m] 0.195 [c] 0.286 [c*m]
	fac4	**0**	**<0.001**	11.241	**<0.001**	7.408	0.099 [m] **0.031 [c]**	13.453	0.106 [m] 0.181 [c] 0.098 [c*m]
	fac5	**0**	**0.01**	13.601	**0.01**	9.835	0.357 [m] **0.028 [c]**	18.104	0.392 [m] 0.195 [c] 0.371 [c*m]

Finally, we also noted a qualitative difference among the colonies. Two females of the group "Woody" in Madrid produced a call type not observed in the other colonies. It was similar to a phee call but was made up of several short intervals, giving the call a distinct rhythm (Figure 2.2). Both animals that performed it also gave a variety of regular phee calls. We only included regular calls in the analysis.

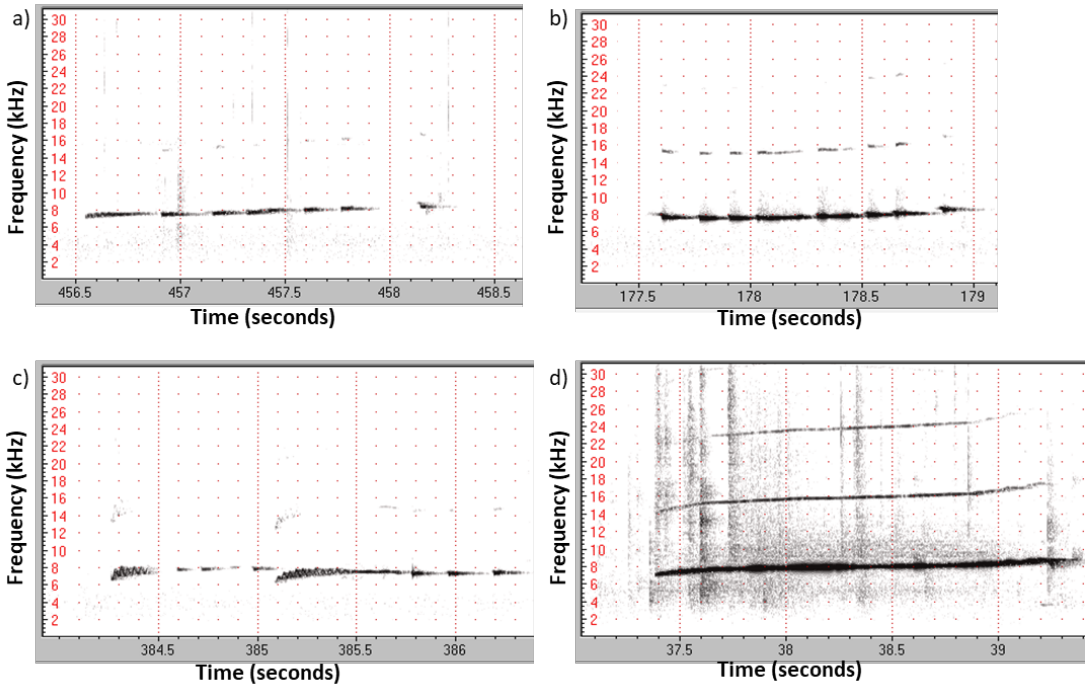

Figure 2.2: Interrupted phee calls (a & b) produced by two common marmoset helper females of the Woody group in the colony at the Universidad Autónoma de Madrid (ES), sampled in October/November 2015. These calls were sometimes produced in combination with a trill call (c). d) represents a regular phee call of one of the two females.

2.5 Discussion

Our results show that three major social calls of common marmosets clearly differed between captive populations, and they did so in 5 (phee and food calls) or 4 (trill Calls) of 5 PCA factors. These differences could not be accounted for by differences in breeding status (i.e., being a breeder or a helper), sex, or body mass. Moreover, two individuals of one family used a call type not found in other groups, suggesting that population differences may go beyond simple accommodation. We do not know when the two females started producing these calls, or whether other animals picked up on these calls over time. Differences in calls can occur in vocal learners if individuals make copying errors while learning novel vocalizations (Slater 1986). These errors could be a source of variation between dialects, and could be maintained over generations if they are passed along with vocal learning.

Our results are consistent with the hypothesis that the marmosets' propensity to engage in vocal accommodation to social partners, both as adults and immatures, leads to population differences in call structure, or dialects [Henry et al., 2015]. Alternatively, however, it is possible that the population differences resulted from differences in environmental factors [Hotchkin and Parks, 2013; Miller et al., 2016]. However, we recorded the animals in Zürich both in indoor and outdoor enclosures, and in rooms with different acoustic properties and they clustered clearly as one population. It is therefore unlikely that environmental differences explain the systematic population differences in all three call types.

To elicit food calls, we gave each animal food we knew it preferred, which differed across the animals. We cannot exclude the possibility that different food stimuli alter the food calls animals produce. However, playback studies with Geoffroy's marmosets (*Callithrix geoffroyi*) suggest they do not [Kitzmann and Caine, 2009]. Presenting the same food to all animals, regardless of whether they liked it or not would been problematic since food calls are typically elicited by highly preferred food only [Vitale et al., 2003]. Food preferences within the colonies were more similar than between colonies. This may in part be explained by the fact that different food types were not fed equally often in all colonies. Additionally, however, it is also consistent with the finding that food preferences can be socially mediated in callitrichids [Snowdon and Boe, 2003]. Still, as we did not test for food preferences specifically and as there are no studies investigating whether common marmosets produce referential food calls that are specific to a food type, we cannot rule out the possibility that some of the variation in food calls can be attributed to the different food types. However, since both trill and phee calls also show population differences, we would expect the same mechanism to act on food calls. Nevertheless, further studies are needed to systematically investigate whether common marmosets encode the food type in their calls or if the differences we found were pure population differences.

The recording settings we used were not always identical, and animals were recorded either alone or in pairs. However, it is unlikely that our results are influenced by these differences. First, we elicited phee calls in all animals by separating them from their group. Second, we recorded food calls alone or in pairs both in Madrid and in Zürich. We did not find separate clusters in the discriminant function analysis for Zürich and Madrid, suggesting that the difference in the recording setup did not influence call production.

Finally, we could not exclude the possibility that genetic drift or founder effects have influenced the animals' morphology and led to the differences in call structure between the colonies. However, body

mass differences did not explain the variation in call structure between the colonies. Although we had only body mass data from two of the colonies (Zurich and Madrid), we are confident that the finding that mass did not influence the parameters we measured is a general finding. YZ estimated the body mass of animals in from Rome to be close to those for Zurich, and the biggest difference to be between Zurich and Madrid. Nevertheless, we cannot rule out other morphological differences across the three colonies. Future studies that directly test whether individuals translocated to a different colony adjust their call structure will be needed to fully rule out genetic influence as a source of population differences.

Together with accumulating evidence for vocal accommodation to social partners [Rukstalis et al., 2003; Snowdon and Elowson, 1999], it is arguably most likely that the population differences we report represent different dialects that occurred due to vocal accommodation and therefore learning within the three colonies. Nevertheless, it remains crucial to further exclude alternative hypotheses by translocating animals between colonies and examining whether they and their offspring accommodate to the new environment and/or the calls of the new colony. Disentangling the role of genetics, environment, and social vocal learning in marmoset vocalizations is fundamental to establish the similarities and differences in the communicative systems of humans and marmosets, as common marmosets are increasingly used as a neuroscientific model for human social behaviour [Miller et al., 2016], including language evolution.

2.6 Corrigendum

Corrigendum

During the further analysis of the previously data used in this paper, we decided not to use the parameter "shimmer" any further, as variation in the amplitude could be mainly influenced by body posture and head tilt. Further, we detected some beforehand unnoticed irregularities in the data. We therefore provide here the results with the corrected data and without shimmer.

Although the numbers in some results change, the **main conclusion of the original paper does not change.**

Table 2.7 cor: Factor loadings of acoustic parameters on five PCA factors (F1-F5) with an Eigenvalue >1 extracted for three call types in common marmosets. Factor loadings larger than + 0.5 or smaller than – 0.5 are marked in bold, indicating that the parameter made a considerable contribution to the factor. For description and definition of the parameters see Table 2. We collected data in February 2014 in the Istituto Superiore di Sanità in Rome, Italy, in 2014 and 2015 at the University of Zürich, Switzerland and in October/November 2015 at the Universidad Autónoma de Madrid, Spain.

	Phee				Trill					Fc				
Parameter	F1	F2	F3	F4	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
sound_duration	0.23	-0.78	0.29	0.10	-0.06	0.58	-0.49	0.38	-0.13	0.18	-0.60	0.17	-0.18	0.21
MeanF0	0.97	-0.01	0.13	-0.08	0.98	0.06	0.02	0.07	0.03	0.98	-0.04	-0.17	-0.01	-0.01
F0start	0.84	-0.01	-0.38	-0.21	0.91	-0.29	0.10	-0.07	-0.08	0.97	0.06	-0.05	-0.14	0.01
F0end	0.84	0.06	0.32	-0.10	0.89	0.20	-0.11	0.21	0.13	0.91	-0.21	-0.29	0.10	0.00
MaxF0	0.92	0.02	0.31	-0.08	0.95	0.25	0.02	0.09	0.01	0.98	0.06	-0.07	-0.10	0.01
TimeMaxF0	-0.15	-0.11	0.66	0.25	-0.28	0.38	-0.34	0.44	0.34	-0.06	0.07	-0.25	0.75	0.12
MinF0	0.88	-0.04	-0.32	-0.17	0.93	-0.31	0.01	0.02	0.03	0.91	-0.21	-0.28	0.08	-0.01
F0absslope	0.07	0.89	0.33	0.04	0.03	0.44	0.69	-0.34	0.08	0.44	0.72	0.22	-0.18	0.10
F0Var	0.05	0.86	0.44	0.00	0.08	0.70	0.47	-0.36	-0.03	0.31	0.74	0.19	-0.34	-0.18
FmRate	-	-	-	-	-0.16	-0.66	-0.07	-0.22	0.02	-	-	-		-
FMextend	-	-	-	-	0.06	0.88	0.18	-0.02	0.00	-	-	-		-
Q25	0.86	-0.05	0.00	0.13	0.50	-0.17	0.22	0.38	-0.32	0.88	-0.10	0.15	0.13	0.02
Q50	0.82	0.09	-0.01	0.40	-0.32	-0.01	0.41	0.61	-0.36	0.55	-0.01	0.65	0.33	0.01
Q75	0.21	0.08	-0.38	0.81	-0.22	-0.14	0.53	0.59	0.19	0.13	-0.09	0.82	0.36	-0.01
Fpeak	0.07	-0.42	0.22	0.03	-0.01	0.28	-0.22	0.14	-0.34	0.07	-0.20	0.03	-0.31	-0.37
Timeofmaxintensity	-0.08	-0.10	0.16	0.47	0.10	0.00	0.01	0.18	0.76	0.02	0.47	-0.31	0.52	-0.44
jitter	0.00	0.50	-0.55	0.08	-0.16	-0.32	0.64	0.21	0.11	-0.01	0.41	-0.15	0.03	0.76
Shimmer	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 2.8 cor: Results of a permutated Discriminant Function analysis (pDFA) testing for population differentiation in common marmoset vocalistations. We obtained expected values by calculating the mean of correctly classified calls in the randomized datasets created in the pDFA. We collected data in February 2014 in the Istituto Superiore di Sanità in Rome, Italy, in 2014 and 2015 at the University of Zürich, Switzerland and in October/November 2015 at the Universidad Autónoma de Madrid, Spain.

Call type	Comparison	n	Correct (%)	Expected (%)	p-value	Crossvalidated correct (%)	Crossvalidated Expected (%)	p-value
Phee	Madrid – Rom – Zürich	1304	65.69	49.32	0.002	62.70	38.67	0.001
	Madrid – Rom	652	82.76	64.77	0.004	88.92	58.13	0.001
	Zürich – Rom	804	80.98	65.22	0.015	79.19	54.01	0.003
	Zürich – Madrid	1152	74.1	59.95	0.002	75.61	55.45	0.001
Trill	Zürich – Madrid	461	86.41	61.83	0.001	84.45	58.41	0.001
Food	Madrid – Rom – Zürich	3598	63.12	45.08	0.001	58.19	40.53	0.001
	Madrid – Rom	2137	81.41	62.34	0.001	77.29	60.90	0.001
	Zürich – Rom	2622	69.29	61.16	0.025	65.66	56.25	0.019
	Zürich – Madrid	2437	77.09	59.63	0.001	74.45	56.38	0.001

Table 2.9 cor: Results of post-hoc analyses testing which Principle Components contributed most to differences in three call types between three populations of common marmosets. Values p<0.05 are marked in bold. (df = degrees of freedom). We collected data in February 2014 in the Istituto Superiore di Sanità in Rome, Italy, in 2014 and 2015 at the University of Zürich, Switzerland and in October/November 2015 at the Universidad Autónoma de Madrid, Spain.

	Phee			**Trill**			**Food**		
	df	f	p	df	f	p	df	f	p
Factor 1	41.934	5.724	**.001**	25.664	14.844	**.001**	38.280	2.260	.118
Factor 2	39.931	4.065	**.025**	26.983	4.978	**.034**	38.191	11.057	**.000**
Factor 3	42.194	1.228	.303	26.027	6.391	**.018**	37.294	19.785	**.000**
Factor 4	41.306	13.801	**.000**	26.620	27.880	**.000**	37.804	7.067	**.002**
Factor 5	-	-	**-**	23.361	.003	.955	36.504	125	.883

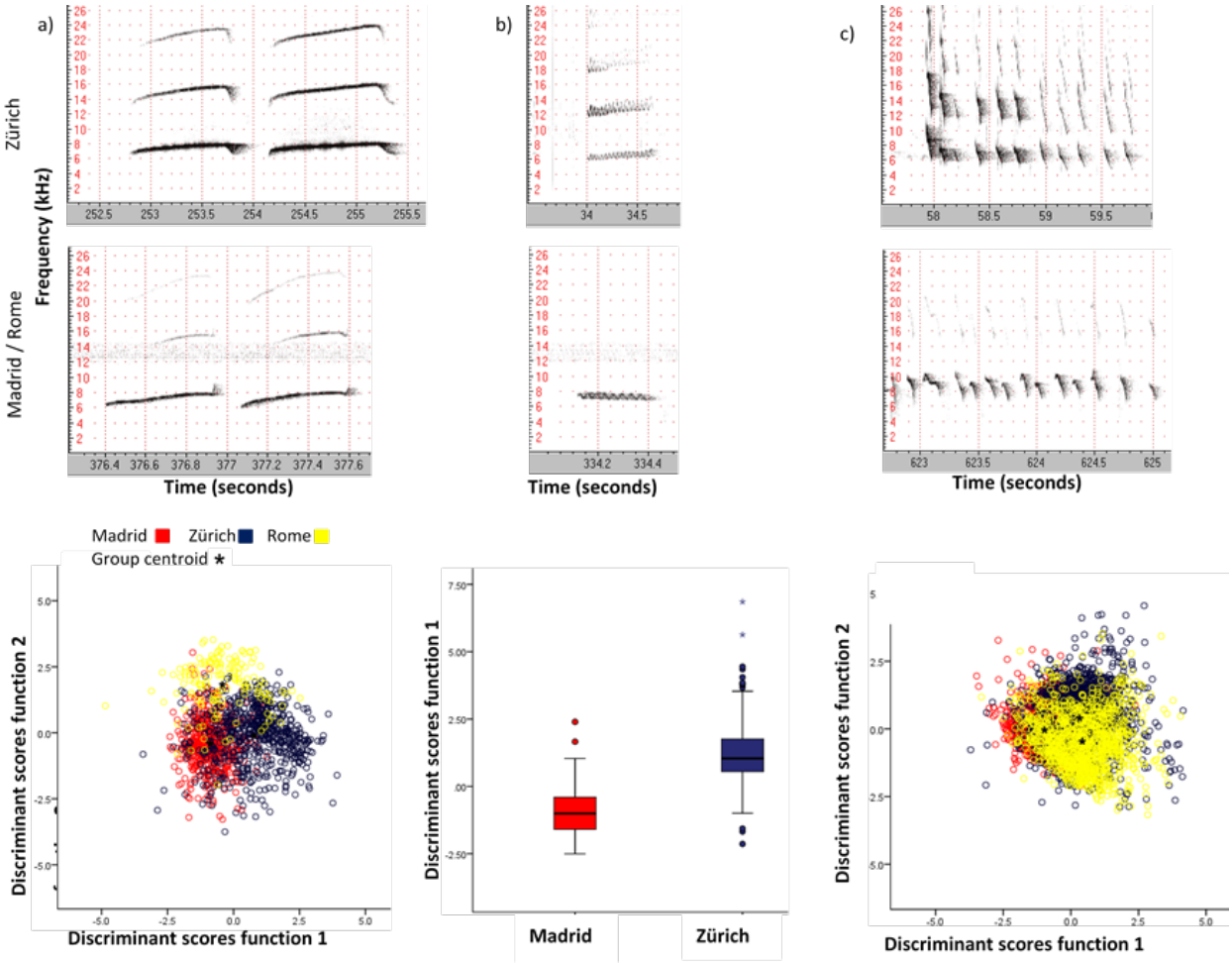

Figure 2.3 corr: **Spectrograms of three call types in common marmosets** (first and second row) **and results of Discriminant Function Analysis to test for population differences** (third row). The first row of the spectrograms shows calls from helper males in Zürich, the second row from a breeding male in Madrid (phee and trill) or from a helper male in Rome (food call). We selected calls close to the group centroid in the discriminant function plots to be most representative. The pDFA results show the first and second discriminant function for a & c and the first discriminant function in b. Populations are: red = Madrid, yellow = Rome, dark blue = Zurich. a) phee calls, p=0.001, b) trill calls, p=0.001 (only Madrid and Zürich due to a too small sample from Rome); c) food calls, p=0.001. We collected data in February 2014 in the Istituto Superiore di Sanità in Rome (Italy), in 2014 and 2015 at the University of Zürich (Switzerland) and in October/November 2015 at the Universidad Autónoma de Madrid (Spain). Boxplots show the median, interquartile range, outliers as circles and extremes as asterisk.

Table 2.10 cor. Result of linear mixed effect models testing for the influence of colony, body mass, colony&body mass, and colony&bodymass on five PCA factors in three different call types in common marmosets. P values < 0.5 are bold, and letters in brackets indicate which factor of the model they belong to ([c] = colony, [m] = mass, [c*m] = interaction between colony and mass). ΔAIC values are calculated as the distance of each model to the respective best model. For description and definition of the parameters see Table 2. We collected data in 2014 and 2015 at the University of Zürich, Switzerland and in October/November 2015 at the Universidad Autónoma de Madrid, Spain.

Call type	**Model**	[colony]		[mass]		[colony], [mass]		[colony], [mass], [colony]*[mass]	
	factor	ΔAIC	p-value	ΔAIC	p-value	ΔAIC	p-value	ΔAIC	p-value
Phee	fac1	0	.548	9.71	.984	9.64	.427 [c] .595 [m]	17.82	.974 [c] .600 [m] .892 [c]*[m]
	fac2	0	.036	10.40	.066	9.33	.241 [c] .518 [m]	16.90	.590 [c] .827 [m] .482 [c]*[m]
	fac3	0	.016	10.62	.031	8.57	.181 [c] .429 [m]	16.70	.175 [c] .204 [m] .235 [c]*[m]
	Fac4	0	.001	11.12	.001	8.32	.058[c] .161 [m]	15.59	.363 [c] .502 [m] .237 [c]*[m]
Trill	fac1	0	.002	15.13	.059	9.17	.017 c] .532 [m]	10.07	.003 [c] .769 [m] .008 [c]*[m]
	fac2	0	.003	7.42	.001	7.46	.429 [c] .115 [m]	15.68	.585 [c] .147 [m] .506 [c]*[m]
	fac3	0	.044	12.16	.254	9.58	.082 [c] .510 [m]	16.30	.274 [c] .391 [m] .173 [c]*[m]
	fac4	0	.000	14.84	.001	9.35	.024 [c] .533 [m]	17.60	.696 [c] .536 [m] .969 [c]*[m]
	Fac5	0	.970	9.40	.904	9.76	.833 [c] .809 [m]	17.77	.542 [c] .876 [m] .558 [c]*[m]
Food	fac1	0	.887	9.27	.943	9.76	.721 [c] .737 [m]	18.20	.982 [c] .736 [m] .943 [c]*[m]
	fac2	0	.139	9.20	.133	9.76	.666 [c] .628 [m]	17.75	.485 [c] .657 [m] .444 [c]*[m]
	fac3	0	.000	19.07	.000	8.12	.001 [c] .134 [m]	16.90	.992 [c] .148 [m] .592 [c]*[m]
	Fac4	0	.362	10.10	.966	8.60	.119 [c] .200 [m]	14.57	.060 [c] .222 [m] .091 [c]*[m]
	Fac5	0	.417	9.44	.494	9.48	.688 [c] .955 [m]	15.76	.138 [c] .840 [m] .143 [c]*[m]

3. Chapter

-

Are dialects socially learned in marmoset monkeys? Evidence from translocation experiments

Are dialects socially learned in marmoset monkeys?

Yvonne Zürcher[1]*, Erik P. Willems[1], Judith M. Burkart[1]

[1]Department of Anthropology, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

*Corresponding author: yvonne.zuercher@uzh.ch

Published in PLOS ONE 14(10): e0222486.

https://doi.org/10.1371/journal.pone.0222486

3.1 Abstract

The acoustic properties of vocalizations in common marmosets differ between populations. These differences may be the result of social vocal learning, but they can also result from environmental or genetic differences between populations. We performed translocation experiments to separately quantify the influence of a change in the physical environment (experiment 1), and a change in the social environment (experiment 2) on the acoustic properties of calls from individual captive common marmosets. If population differences were due to genetic differences, we expected no change in the vocalizations of the translocated marmosets. If differences were due to environmental factors, we expected vocalizations to permanently change contingent with environmental changes. If social learning was involved, we expected that the vocalizations of animals translocated to a new population with a different dialect would become more similar to the new population. In experiment 1, we translocated marmosets to a different physical environment without changing the social composition of the groups or their neighbours. Immediately after the translocation to the new facility, one out of three call types showed a significant change in call structure, but 5-6 weeks later, the calls were no longer different from before the translocation. Thus, the novel physical environment did not induce long lasting changes in the vocalizations of the marmosets. In experiment 2, we translocated marmosets to a new population with a different dialect. Importantly, our previous work had shown that these two populations differed significantly in vocalization structure. The translocated marmosets were still housed in their original social group, but after translocation they were surrounded by the vocalizations from neighbouring groups of the new population. The vocal distance between the translocated individuals and the new population decreased for two out of three call types over 16 weeks. Thus, even without direct social contact or interaction, the vocalizations of the translocated animals converged towards the new population, indicating that common marmosets can modify their calls due to acoustic input from conspecifics alone, via crowd vocal learning. To our knowledge, this is the first study able to distinguish between different explanations for vocal dialects as well as to show crowd vocal learning in a primate species.

Keywords: vocal learning, common marmoset, primates, social accommodation, translocation experiments, crowd vocal learning

3.2 Introduction

Population differences in vocalizations have been reported for many bird species [Henry et al., 2015], but also for mammals including primates [Lameira et al., 2010]. Such population differences can be the result of environmental differences, if animals adapt to a vocal optimum in their local environment to increase signal transmission (i.e. environmental accommodation [Ruch et al., 2018]). Alternatively, they can be the result of genetic differences, for instance when the latter lead to population level differences in vocal tract morphology. Finally, population differences in vocalizations can be the result of cultural transmission, i.e. vocal social learning or social accommodation. These latter cases are of particular interest because of their similarity with human dialects and potential implications for language evolution [Fedurek and Slocombe, 2011; Lameira et al., 2010; Ruch et al., 2018]. The terminology in vocal population differences is somewhat ambiguous in the animal literature. In humans, the word "dialect" is mainly used for lexical or grammatical differences, but according to Wolfram [Wolfram, 1998] it can include any language differences including pronunciation as well as language use. In birds, the term dialect was mainly used for vocal differences between neighbouring colonies that could interbreed, while differences between isolated populations were often termed "geographic variation", although the two terms can be understood as two points on a continuum [Baker and Cunningham, 1985]. In primates, finally, dialect is often used for any acoustic differences between populations, including structural differences of the same calls (comparable to pronunciation differences in human dialects) (e.g. [de la Torre and Snowdon, 2009; Mitani et al., 1992; Schlenker et al., 2014]). Here, we will follow the tradition of the primate literature and refer to population differences in acoustical call structure as dialects, following the definitions of [Henry et al., 2015; Wolfram, 1998], but see [Lameira et al., 2010] for a use of terminology more consistent with human literature.

Dialects in many birds are the result of cultural transmission, which is consistent with their propensity for vocal learning [Henry et al., 2015]. In contrast, in primates, vocal learning is rare [Egnor and Hauser, 2004] and current evidence that social factors play a role in population differences in vocal structure is indirect [Lameira et al., 2010]. For instance, it has been argued that since vocal distance between chimpanzee populations is not correlated with geographical distance, genetic explanations are unlikely [Crockford et al., 2004]. Likewise, available evidence supporting a role of vocal learning during ontogeny in callitrichid monkeys [Elowson et al., 1998a; Elowson et al., 1998b; Gultekin and Hage, 2017; Gultekin and Hage, 2018; Snowdon and Elowson, 2001; Takahashi et al., 2016; Takahashi et al., 2015] may imply that population differences in these species are the result of vocal learning. However, whereas this evidence is certainly consistent with the idea that social accommodation is responsible for the emergence of population differences, we nevertheless cannot automatically assume that such ontogenetic effects are sufficient. Furthermore, environmental factors clearly can affect vocal structure in primates [Ruch et al., 2018], including callitrichids [de la Torre and Snowdon, 2002; de la Torre and Snowdon, 2009], and therefore are a likely source for population differences too [de la Torre and Snowdon, 2009; Tanaka et al., 2006]. Population differences in vocal structure may thus well be the result of mechanisms other than social vocal learning in these species.

The gold standard for identifying the origin of population differences in behaviour, including vocalizations, is the use of translocation experiments [Laland and Hoppitt, 2003]. These experiments allow to test whether population differences in vocalizations merely reflect genetic differences, as well as to separate

the effect of environmental factors (by moving animals to a new location without changes in social constellation) and of social factors (by moving animals to a new population with a different dialect).

Our goal was to use translocation experiments to investigate the origin of population differences in common marmosets (*Callithrix jacchus*). Marmosets are callitrichid monkeys that show a high level of vocal plasticity [Snowdon, 2013; Snowdon, 2017] and well-established population differences in vocalizations in the wild (pygmy marmosets (*Cebuella pygmaea*): [de la Torre and Snowdon, 2009]) and in captivity (common marmosets: [Zürcher and Burkart, 2017]). We performed two translocation experiments with animals from two captive populations with known dialects [Zürcher and Burkart, 2017] (experiment 1 and 2). If population differences were the result of genetic differences only, we predicted no change in vocalizations in response to any translocation. If they were the result of environmental accommodation, we expected their vocalizations to change permanently in a novel environment (experiment 1). If population differences were the result of social accommodation, we expected that animals translocated into a new population with a different dialect (experiment 2) would accommodate their vocalizations and become more similar to the vocalizations of the new population.

In experiment 1, we recorded three types of vocalizations of marmosets after translocation to a new facility: trill calls, phee calls and food calls. Trill calls are short distance contact calls, phee calls are long distance contact calls, and food calls are mainly produced in the context of discovering and sharing preferred food [Bezerra and Souto, 2008; Vitale et al., 2003]. We found that immediately after the translocation, food calls showed a significant change in call structure, but 5-6 weeks later, the calls were no longer different from before the translocation. Thus, the novel environment did not induce long lasting and permanent changes in the vocalizations. Consistent with these results, we also did not find any change in calls in an additional set of four individuals, 3-7 weeks after they had been translocated to a new physical environment (from their natal colony to a quarantine station without neighbours, see experiment 2).

In experiment 2, we translocated focal individuals from their original population via a quarantine station to a new target population. Importantly, the original and the target population show known vocal differences in trill calls, phee calls, and food calls, meaning that animals of each colony were acoustically more similar to other animals of their colony than to animals of other colonies [Zürcher and Burkart, 2017]. We developed a statistical procedure to express and quantify vocal distances between the focal individuals and status- and sex-matched individuals from the target population *before* and *after* the translocation. We found that over a period of 16 weeks after translocation, the focal individuals were becoming more similar to the target population baseline with regard to trills and phee calls, and more dissimilar in food calls. Since in the new colony the focal individuals were still housed in the same social group as in the home colony and were exposed only passively to the vocalizations of the individuals from the target colony, this effect corresponds to crowd vocal learning recently reported in bats [Prat et al., 2017].

Together, the results of the translocation experiments revealed that marmoset calls are not strictly genetically fixed and that population differences are most likely the result of social vocal learning. Social vocal learning in common marmosets can thus occur by passive exposure to a different dialect, without direct social interaction, via crowd vocal learning.

3.3 Results

3.3.1 Experiment 1: Environmental accommodation

To test whether common marmosets vocally accommodate to their environment, we recorded the vocalizations of eight common marmosets (referred to as "ZH animals", a subset of the ZH colony) before and after they were translocated to a new building (a new acoustic environment, but the same social environment, "translocation A" in Fig 1). We recorded the animals regularly before and after the translocation event: immediately before and after the move (recordings "*Before"* and "*After1*" in Figure 3.1) and again 5 – 7 weeks later ("*After2*" in Figure 3.1).

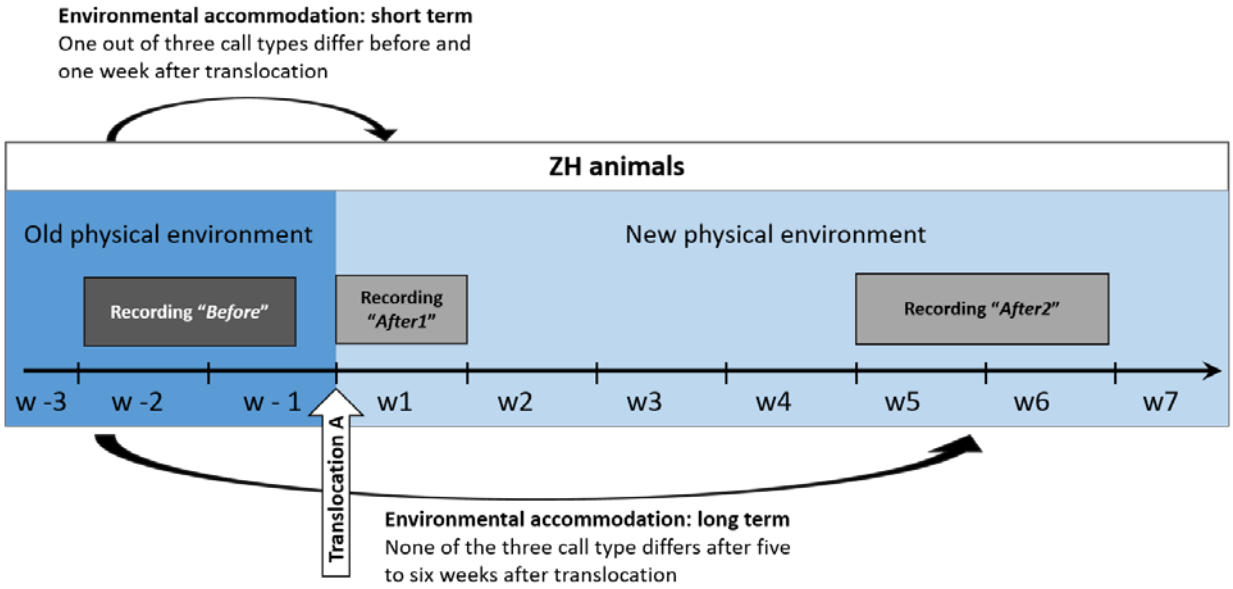

Figure 3.1: Timeline and comparison of calls before and after translocation A. The vocalizations of eight common marmosets were recorded in their familiar surrounding (old physical environment) over two weeks ("Before"). After this, the whole colony was moved to a new location. As soon as the eight monkeys were translocated to the new physical environment, they were recorded again during one week ("After1"), as well as after five to six weeks ("After2"). To investigate if changes in the physical environment lead to short-term environmental accommodation, we compared the recordings "Before" to the recordings "After1" (upper black arrow). We found a significant change in food calls and a strong, but non-significant trend in phee calls, but no change in trill calls. To further analyses if these differences were stable over time (long-term environmental accommodation), we compared the recordings "Before" to the recordings "After2" (lower black arrow). None of the call types was different from before the translocation anymore, suggesting that the new environment did not lead to long-term accommodation.

We recorded three different call types (trill calls, phee calls, and food calls) and extracted 15 (phee - and food calls) or 17 (trill calls) parameters for each call, using the program Praat [Boersma and Weenink, 2009]. We did not include compound calls that show characteristics of two different call types, like trillphees. To reduce the number of parameters included in the analysis, we performed a principal component analysis (PCA) for each call type, resulting in 3 - 4 PC-Factors per call type, explaining 64.45%, 71.68% and 72.84% of the total variation respectively (see supplementary table Table S3.5 for factor loadings). First, we tested whether the calls given before and immediately after the translocation could be distinguished. We performed a crossed permutated Discriminant Function Analysis (pDFA) [Mundry and Sommer, 2007] and found that food calls differed significantly before vs. immediately after the

translocation, but trill calls did not. Phee calls did not differ significantly, but showed a strong trend (Table 3.1).

Table 3.1: Short-term environmental accommodation. In a sample of eight common marmosets from the ZH population, one of three call types showed a significant difference and one a strong trend to differ in the acoustic properties in the week immediately after translocation, when compared to the calls before the translocation. % of expected calls indicate the amount of correct classification if classification were random, % actual correct the amount of correct classification the pDFA reached with the data. Significant p-values are indicated in bold, trends in italics.

	Call type	Number of animals	Number of calls	% expected correct	% actual correct	p-value
ZH animals "Before" – "After1"	Trill call	8	311	56.28	59.59	0.325
	Phee call	8	607	54.09	60.30	*0.052*
	Food call	8	1292	56.23	68.51	**0.028**

To assess whether these changes were permanent, or perhaps only an artefact of increased stress levels due to the move, we recorded the same animals again 5 – 6 weeks after the translocation event ("*After2")*. If the animals actually accommodated to the new physical environment, we predicted that the changes in calls should remain stable over time or become stronger, if accommodation takes more time than one week. We again performed a pDFA to test if the calls were still different from the calls before the move. We found that the changes observed immediately after the translocation did not persist over time, and calls from the period "*After2"* were no longer different from the period "*Before"* (Table 3.2). We thus did not find evidence for environmental accommodation after the translocation to the new facilities.

We corroborated this pattern by also analyzing four additional individuals that were translocated from another colony in Madrid, Spain (referred to as "MA animals") to a quarantine station, before they were integrated into the ZH colony (see below, Fig 2 and experiment 2). We recorded these animals both in their natal group ("*Before*") and later on in the quarantine station ("*NewPhys*"). In the quarantine station, they were housed with a same-sex sibling from their natal group, thus in a new environment but without novel vocal input. We recorded the MA animals 4 – 7 weeks after arriving in the quarantine, i.e. after potential stress from the travel should have abated. We proceeded identically as with the calls from the ZH animals by first performing a PCA (resulting in 4 factors each, which in total explain 72.24%, 76.67% and 70.65% of the total variation respectively, see supplementary Table S3.6 for factor loadings) and then a crossed pDFA to test whether the calls differed before and after the translocation. We did not find any structural differences in the calls of the MA animals for any of the three call types (Table 3.2), although the physical structures of the two recording facilities were substantially different (mainly a small wooden hut in Madrid vs a large concrete room in the quarantine). It therefore seems that translocating common marmosets to a physically different captive environment did not have any long lasting effects on their call structure.

Table 3.2: Long-term environmental accommodation. ZH animals that were recorded again 5 – 6 weeks after translocation ("After2"), as well as four additional MA animals that were recorded 4 – 7 weeks after translocation ("NewPhys"), did not differ in their call structure compared to before the translocation. The new environment thus did not lead to persistent, long-term environmental accommodation. % of expected calls indicate the amount of correct classification if classification were random, % actual correct the amount of correct classification the pDFA achieved with the data.

	Call type	Number of animals	Number of calls	% expected correct	% actual correct	p-value
ZH animals "Before" - "After2"	Trill call	8	279	52.63	59.41	0.107
	Phee call	8	493	56.12	62.43	0.122
	Food call	8	895	52.77	54.62	0.402
MA animals "Before" – "NewPhys"	Trill call	4	278	64.29	73.31	0.389
	Phee call	4	833	57.05	64.05	0.156
	Food call	3	320	64.08	83.61	0.121

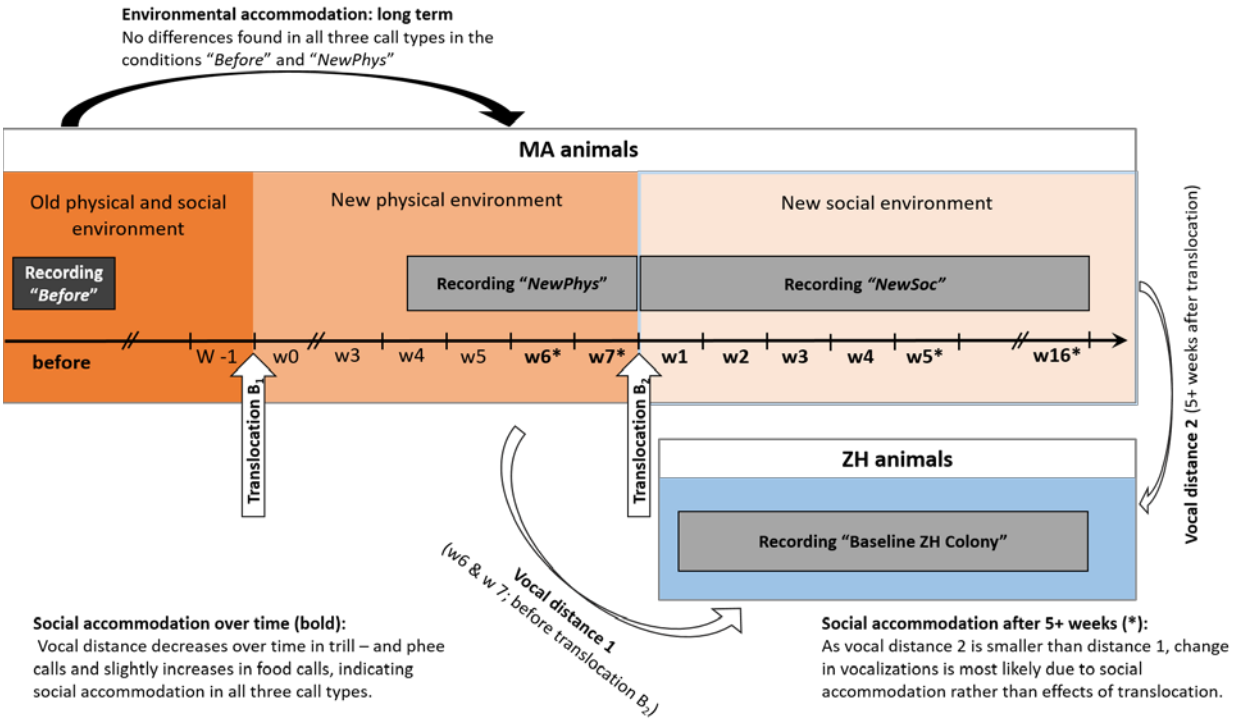

Figure 3.2: Timeline and comparisons of translocation B_1 and B_2. The vocalizations of the animals from MA were first recorded in their home colony (old physical and social environment, recording "Before"). After that, they were translocated to a new physical environment ("Translocation B_1", to a quarantine station, recording "NewPhys"), which allowed us to re-assess **long-term environmental accommodation** (black arrow). Their calls did not differ before and after the translocation B_1 to the new physical environment. Next, the animals were translocated to the ZH colony ("Translocation B_2", to a new social (and physical) environment, recording "NewSoc"). To quantify the **social accommodation over time**, we measured the vocal distance before any new social vocal contacts, during the weeks 6 & 7 after translocation B_1, as well as immediately after the animals arrived in the new colony and up to week 16, (weeks labelled in bold). To test for **social accommodation after 5+ weeks** of contact with the new colony (when potential short-term environmental accommodation effects must have disappeared), we compared the vocal distance (white arrows) between the animals from the ZH colony ("Baseline ZH Colony") and the MA animals,

immediately before Translocation B_2 (vocal distance 1) and after they were 5+ weeks in the new social environment (after "Translocation B_2"; vocal distance 2, weeks with *). We found that after translocation to the new social environment, the vocal distance between the MA and the ZH animals decreased in two call types and slightly increased in the third call type, as well as remained smaller / larger also after 5+ weeks, which is indicative of social accommodation.

3.3.2 Experiment 2: Social accommodation

To test whether common marmosets showed social vocal accommodation after passive exposure to a new colony, i.e. crowd vocal learning, we translocated the four MA animals to the ZH colony (after they had spent seven weeks in quarantine, see Fig 2, translocation B_1 and B_2). In a previous study we could show that calls from animals of the MA population and the ZH population are significantly distinct [Zürcher and Burkart, 2017]. It is important to note that this vocal differences between colonies indicates that animals from one colony are generally most similar to other animals of their colony but differ from animals of another colony. Animals of one colony share a dialect with one another, and by this also share typical acoustic features that make the colonies distinct.

After translocation B_2, the translocated animals were still housed as sibling pairs, but in the same colony room as the ZH animals. This allowed for vocal and perhaps olfactory contact, but not for direct physical interaction or visual contact. We collected vocal recordings of the four MA animals as well as four status- and sex-matched ZH colony members ("*Baseline ZH colony*"). To quantify **vocal accommodation over time**, we measured vocal distance before each translocation as well as each week after translocation B_2 for up to 16 weeks (see Fig 2).

First, we performed a PCA which resulted in 4 factors being retained, cumulatively accounting for 66.6.0%, 75.1%, and 70.8% of the total variation for trill calls, phee calls and food calls, respectively (see Supplementary Table S3.7 for factor loadings). Using these factors, we calculated the Euclidian distance between each measured call and its corresponding average from the four ZH animals. This resulted in several distance values per animal, per call type, and per week. To test how vocal distance developed over time in different call types, we performed a series of different linear models (see methods for details on model selection). The best model (Table 3.3) for trill calls revealed an interaction effect of sex and time, indicating that males showed stronger convergent accommodation than females. The best model for phee calls revealed a significant effect of both sex and time, whereas for food calls, it indicated a small but significant effect of time in the opposite direction (divergence) in food calls.

Overall, the results show that for trill calls and phee calls, the vocal distance between the ZH and MA animals decreased over time, whereas it slightly increased in food calls. As the ZH baseline value was the same for all points in time in this analysis, the observed change in distance was most likely induced by the MA animals changing their vocalizations (see Table 3.3 and Figure 3.3).

Table 3.3: Social accommodation over time. The effect of sex, exposure time and the interaction thereof on vocal distance (ln-transformed Euclidean distance) between the vocalizations of four translocated MA individuals and the call type specific average of the new population ("Baseline ZH colony") for each call type. Parameter estimates, standard errors, and statistical significance are obtained from a linear mixed effects model. Significant (highest-order) effects are indicated with p-values in bold, trends in italics.

Trill

	B	**SE**	**t**	**P**
Intercept	0.062	0.07		
Sex				
Female vs. Male	-0.137	0.088	-1.544	0.177
Exposure time	-0.052	0.008	-6.197	< 0.001
Sex * Exposure time				
Female vs Male interaction with week	0.032	0.009	3.494	**< 0.001**
$N_{obs.}$= 515 from 4 individuals; R^2_m= 0.114, R^2_c= 0.146 χ^2_{MLT}=48.209, p< 0.05				

Phee

	B	**SE**	**t**	**P**
Intercept	-0.008	0.062		
Sex				
Female vs. Male	-0.324	0.085	-3.801	0.023
Exposure time	-0.021	0.005	-3.932	**0.02**
$N_{obs.}$= 1852 from 4 individuals; R^2_m= 0.102, R^2_c= 0.142, χ^2_{MLT}= 10.2724, p< 0.05				

Food

	B	**SE**	**t**	**P**
Intercept	-0.531	0.034		
Sex	0.100	0.046	2.160	0.151
Female vs. Male				
Exposure time	0.009	0.003	2.829	**0.005**
$N_{obs.}$= 1759 from 4 individuals; R^2_m= 0.009, R^2_c= 0.0114, χ^2_{MLT}= 9.846, p< 0.05				

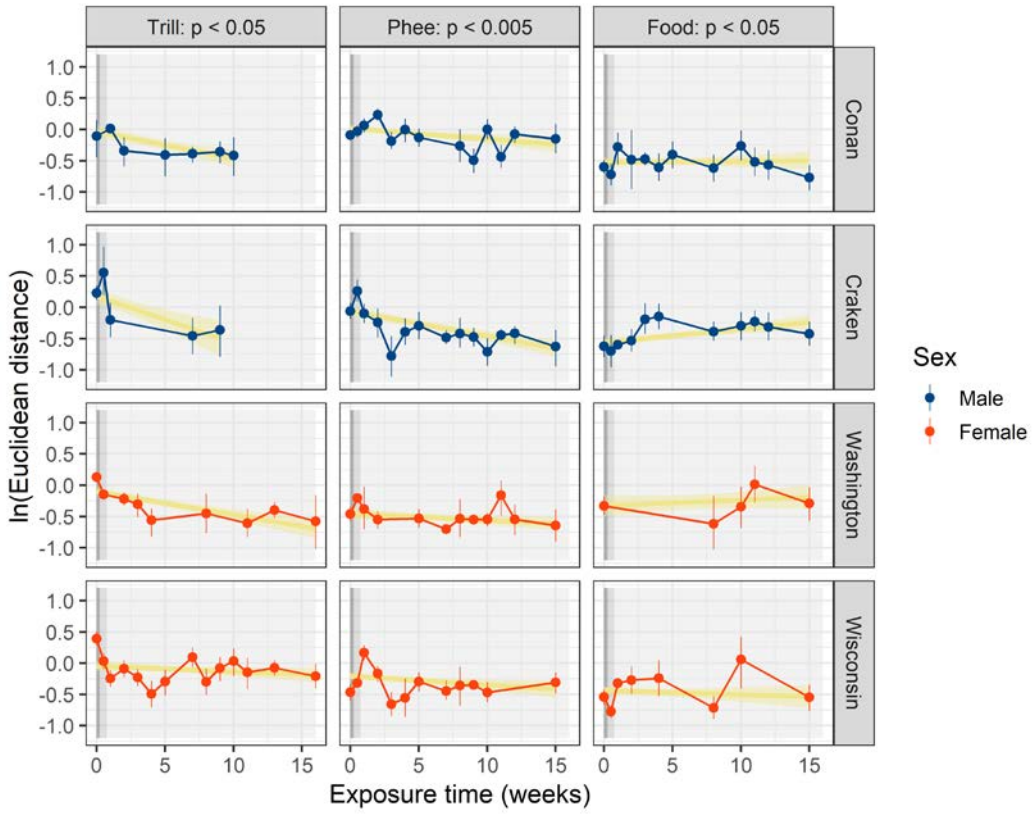

Figure 3.3: Social accommodation over time Changes in vocal distance over time relative to the ZH population for the four translocated animals (rows), separately for each call type (columns). Dots represent weekly averaged values, calculated on a minimum of five recordings per individual per call type, with associated 95% (bootstrapped) confidence intervals represented by vertical error lines. Solid dark-yellow lines with shaded 95% confidence intervals represent the general pattern in change over time. Vocal distance decreases significantly over time in trill- and phee calls, and slightly increases in food calls.

Since the animals were at the same time also translocated to a novel environment (from quarantine to ZH colony room), some change in vocal structure may be due to short-term environmental accommodation, as we had found in experiment 1. We therefore also analysed **social accommodation after 5+ weeks**, i.e. when potential short-term environmental effects should have had disappeared.

We thus compared whether the vocal distance from the MA animals to the ZH baseline immediately before translocation (i.e. the last two weeks in the *"NewPhys"* condition, vocal distance 1 in Figure 3.2) was significantly larger than after 5+ weeks in the new colony (vocal distance 2 in Figure 3.2). We found that in both trill calls and phee calls, vocal distance was significantly smaller after the animals had spent at least 5 weeks in the new colony, whereas in food calls, the vocal distance increased slightly but significantly after 5+ weeks (see Figure 3.4 and supplementary Table S3.8 for the models). Social accommodation effects thus persisted over time and were not an artefact of short-term environmental accommodation.

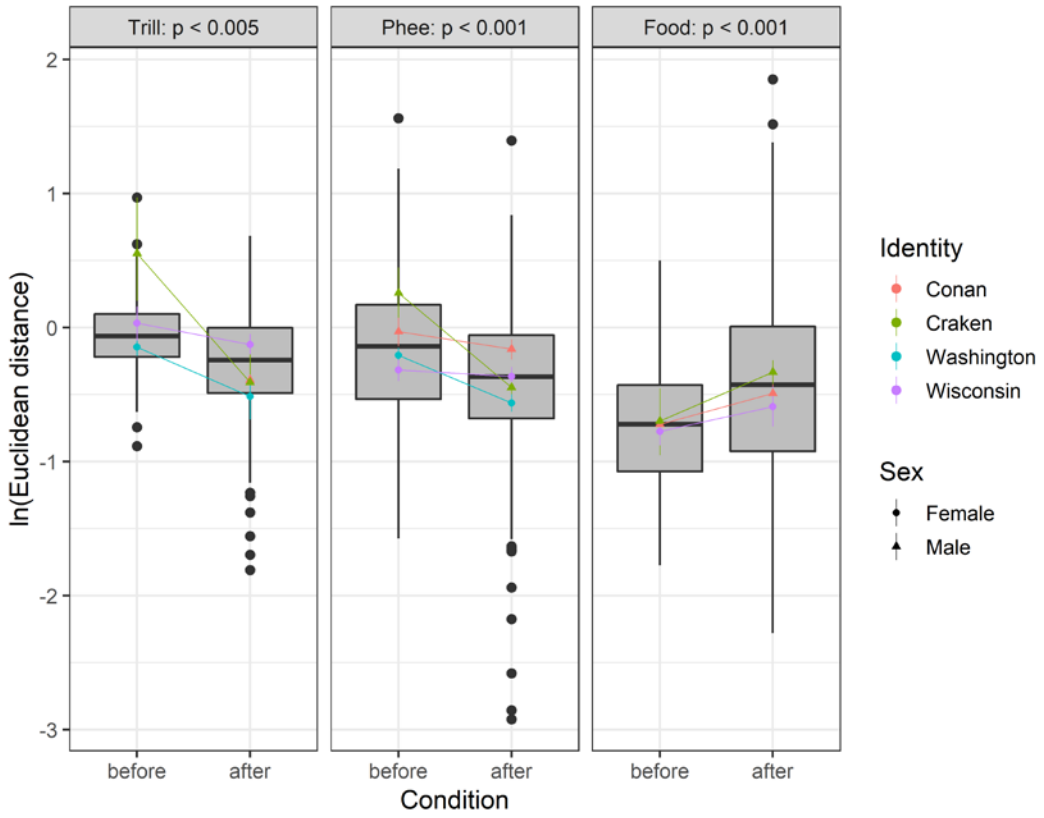

Figure 3.4*:* Social accommodation after 5+ weeks. The vocal distance was significantly larger immediately before translocation (last two weeks of the "NewPhys" condition = vocal distance 1 in Fig 2) than after 5+ weeks in the new colony (vocal distance 2 in Fig 2) for both phee- and trill calls. Food calls showed a slightly larger vocal distance after translocation, but effect size is small. Social accommodation thus persists over time and cannot be explained by short-term environmental accommodation.

As we tested changes in vocal distance in call structure based on the combination of all call parameters, it is not straightforward to identify which parameter contributed most to the change in vocal distance. We nevertheless provide the average parameter for each individual and each call type before and after the translocation in the Supplementary Table S3.9 - Table S3.12 for an overview, but emphasize that it is the combination of all rather than a single parameter that was responsible for the observed results.

3.4 Discussion

Vocal dialects have been reported in marmosets in the wild [de la Torre and Snowdon, 2009] and in captivity [Zürcher and Burkart, 2017], but to date it was not clear whether these population differences are the result of genetic differences, of environmental accommodation or of social vocal learning. Translocation experiments, as the ones reported here, provide an ultimate test to discriminate between these possibilities. In these experiments, individuals with their social partners are translocated to a different environment to test for environmental influences, or to a different population with a different dialect to test for social influences. The vocalizations of the individuals are constantly recorded, to allow a comparison of the call structures before and after translocation. To our knowledge, such studies are rare in the primate literature so far. Previous translocation studies found parallel changes, rather than

divergence or convergence, in the vocalizations of pygmy marmosets after two social groups got in acoustic contact with each other [Elowson and Snowdon, 1994]. Likewise, Rukstalis et al. could show that vocalizations of Wied's black tufted-ear marmosets (*Callithrix kuhlii*) modified their phee calls after acoustic contact with unfamiliar individuals, although their analysis did not allow to quantify the direction or amount of change in the call properties [Rukstalis et al., 2003]. The only study using a translocation approach in apes was done with chimpanzees, but unfortunately, crucial data of the subjects' vocalizations before the translocation event was missing, and the study therefore had only limited potential to answer the question of vocal production learning in chimpanzees [Fischer et al., 2015; Watson et al., 2015a; Watson et al., 2015b]. The current study is adding to this research by measuring the actual amount of accommodation, using a newly developed tool to quantify vocal distance. The results show that vocal dialects in common marmosets are not the result of genetic differences or environmental accommodation but arise due to social vocal learning.

Evidence for environmental accommodation after translocation was limited. We did find changes in call structure immediately after the translocation to a novel environment in one out of three call types, but these changes were not permanent and after five weeks, the calls were no longer different compared to before the translocation. This finding was corroborated in a second set of animals, where the change in environment was even larger (a small wooden hut or outdoor cage vs a large concrete room) yet no long-term effects could be detected. Thus, the environmental differences between the MA colony and the ZH colony could not account for the different dialects of these colonies. A likely explanation is that the changes documented during the first week after the translocation were due to the unfamiliar situation and potentially increased stress levels, rather than representing an adaptation to the acoustic properties of the new room. Nevertheless, environmental accommodation has been shown in callitrichid monkeys, including marmosets [de la Torre and Snowdon, 2002; Snowdon and de la Torre, 2002]. Presumably, these environmental differences were more dramatic compared to the ones at stake in the present study, as environmental noise like wind and water (rivers, rainfall), as well as calls from other animals were much more prominent (and different) in the environments of wild populations [de la Torre and Snowdon, 2002] and mostly absent in our captive populations. It may thus well be that differences in acoustic properties in wild habitats can account for some amount of vocal differences, as has been shown for macaques and baboons [Ey et al., 2009; Tanaka et al., 2006]. What our study suggests, however, is that such environmental factors are not necessary for the emergence of dialects in marmosets.

To test whether marmosets could socially learn a new vocal dialect, we translocated individuals between two colonies (MA and ZH) with different vocal dialects [Zürcher and Burkart, 2017]. The vocal distance from the MA animals to the ZH baseline decreased over time after the MA animals had been translocated to the ZH colony, for two out of three call types. Social accommodation showed a slightly different pattern between males and females, and tended to be stronger in the former. However, as the sample size was rather small (two males and two females each), and the two same-sex animals were full siblings, we cannot conclude whether the observed difference is due to a general differences between the two sexes or whether it could be explained by family differences.

The second translocation included a change in social environment, but also in physical environment. Thus, short-term environmental accommodation may have played a role too. To exclude short-term

environmental effects, we corroborated the results of social accommodation by comparing the calls immediately before translocation (*"NewPhys"*) with only the calls that were produced after 5 weeks or more in the new colony (when short-term environmental effects have disappeared, see experiment 1 on environmental accommodation). These additional analyses confirmed the presence of social accommodation: the vocal distance was again significantly smaller after 5+ weeks both in trill calls as well as in phee calls, and the change in vocal distance was again larger in males than in females (see Supplementary Table S3.8). In food calls, we found a small but significant increase in vocal distance. Food calls are mainly produced in the context where animals are willing to share food. As the MA subjects did not have any social bond to any of the ZH animals, and their only sharing partner was their sibling (with the same dialect), it could be that this divergence conserved the group identity and made it more distinct whom the caller was addressing.

In this study, we only investigated the vocal accommodation potential of the animals that were translocated (the MA animals). It may well be that the ZH individuals in acoustic contact with the new MA animals also changed their vocalizations in the direction of the MA animals. Unfortunately, we did not have enough recordings from the ZH animals to analyse these potential changes systematically as well. Note, however, that whether the ZH animals also changed their vocalizations or not is irrelevant for our conclusions, as the change of the MA animals towards the ZH animals per se shows that changes in vocalizations due to a social template is possible and occurs.

As marmoset vocalizations seem to be flexible to a certain degree and can change due to novel social inputs, we can also exclude genetic differences as a potential source of population differences, which suggests that social accommodation is thus the most likely explanation for the different dialects.

The effects found in this study were moderate. However, it is important to keep in mind that social accommodation in trill calls and phee calls occurred even though the MA animals were still housed with their siblings only, and were merely exposed to the vocalizations in the new colony. Learning a dialect by simply being exposed to it, without any direct interaction with animals of the new dialect, has so far, for nonhuman mammals, only been shown in bats [Prat et al., 2017]. Our results show that common marmosets too can learn a vocal dialect by simply being exposed to it, and thus show what Prat et al. call crowd vocal learning [Prat et al., 2017]. Impressively, common marmosets show this skill even as fully mature adults, and not in the developmental phase as it was the case in the bats. To fully understand the range of vocal accommodation in marmosets, it will be necessary to also study it when individuals are not only passively, but also actively exposed to the new dialect, i.e. when they are newly housed with animals with a different dialect and can have direct social contact with each other.

Overall, our results are in line with an increasing body of evidence suggesting some vocal learning potential in common marmosets. Both during ontogeny and as adults, marmosets modify their vocalizations due to social feedback. During ontogeny, exposure to calls of a caregiver plays an important role in the acquisition of vocalizations in immature marmosets [Takahashi et al., 2015]. Contrary to e.g. infant squirrel monkeys or macaques [Hammerschmidt et al., 2001; Owren et al., 1992], common marmosets with interrupted auditory feedback from caregivers did not make the full transition from a juvenile to an adult call repertoire, suggesting that vocal feedback is essential for these primates to produce correct vocalizations [Gultekin and Hage, 2017; Takahashi et al., 2017]. Furthermore, similar to

human infants, marmoset infants show an extensive babbling phase during their development [Elowson et al., 1998a; Elowson et al., 1998b]. In pygmy marmosets, higher amounts of babbling leads to faster acquisition of adult-like calls [Snowdon and Elowson, 2001], suggesting that babbling in marmosets might serve as practice for later vocal production just as in humans [Snowdon, 2001]. Intriguingly, immatures not only adjust their vocal development to parental feedback, but increasing evidence suggests that callitrichid parents themselves adjust their vocal feedback contingent on infant vocalizations [Chow et al., 2015; Takahashi et al., 2015]. Adult marmosets show considerable flexibility in their call production too. They take part in turn taking exchanges, which requires precise coordination of their calls with their communication partner [Takahashi et al., 2013], and plan the timing of their calls to avoid predictable intervals of interfering background noise [Roy et al., 2011]. Finally, several species of the callitrichidae-family have been shown to modify their calls due to changes in their social environment, like new group members [Elowson and Snowdon, 1994; Rukstalis et al., 2003] or new breeding partners [Snowdon and Elowson, 1999]. Together, this body of evidence is fully consistent with our finding that genetic differences are not sufficient to account for differences in marmoset vocal structure, and that they even engage in crowd vocal learning. At the same time, it strongly suggests that immediate social context is a crucial factor for marmoset vocal learning. Accordingly, the effects of crowd vocal learning (convergence) were moderate, and only present in two out of three call types. Based on the recent literature, we hypothesise that social vocal learning effects will be much more prominent once the MA animals will be fully integrated in social groups with ZH animals, allowing for direct social contact.

Marmoset monkeys are increasingly used as a model species for neuroscientists to study the neuronal basis of behaviour, including vocal behaviour [Eliades and Miller, 2017; Marx, 2016; Miller et al., 2016], in order to gain a better understanding of the evolutionary origin of human language. A detailed understanding of their vocalization behaviour, gained by careful observation and experimental manipulations, is a necessary precondition for this endeavour, and this study contributes to providing the necessary and solid base upon which neuroscientific studies can be built on [Marshall-Pescini et al., 2016]. Such studies continue to reveal that marmosets share a surprising amount of features with human language, and an intriguing hypothesis is that this may be linked to their social system. Like humans, marmosets are cooperative breeders, which might have shaped both their need as well as their social mind-set to favour vocal communication and coordination, in a similar way as it may have occurred during human evolution [Burkart et al., 2018].

3.5 Methods

3.5.1 Subjects

We recorded at total of 10267 calls from 16 common marmosets for this study (4337 calls for environmental accommodation and 5930 calls for social accommodation; see Table 3.4). All animals were housed in pairs with either a sibling or a breeding partner in cages structured with a variety of wooden branches, ropes, tubes and other enriching material. The monkeys received food twice a day (vitamin enriched mush in the morning, mixed fruits and vegetables around midday) and various kind of animal proteins and insects and / or gum several times the day. Water was available ad libitum. The experiments were approved by the Kantonales Veterinärsamt Zürich, licence number ZH223/16.

Table 3.4: Overview of the number of calls used in the different analyses and of the individual contribution to the sample.

Translocation	ID	Origin	Sex	Trill calls	Phee calls	Food calls
A	Kaliper	ZH	M	87	67	178
A	Kapi	ZH	M	85	119	173
A	Kapo	ZH	M	51	121	105
A	Marlene	ZH	F	57	62	282
A	Tabor	ZH	M	22	166	348
A	Thilo	ZH	M	32	167	203
A	Vesta	ZH	F	59	26	117
A	Vito	ZH	M	39	80	183
	Total			**432**	**808**	**1589**
B_1	Conan	MA	M	18	298	128
B_1	Craken	MA	M	29	102	43
B_1	Washington	MA	F	136	128	0
B_1	Wisconsin	MA	F	95	305	149
	Total			**278**	**833**	**320**
B_2	Conan	MA	M	114	557	501
B_2	Craken	MA	M	45	315	545
B_2	Washington	MA	F	113	298	149
B_2	Wisconsin	MA	F	168	393	443
ZH Baseline	Gatto	ZH	M	16	304	555
ZH Baseline	Nautilus	ZH	M	33	164	76
ZH Baseline	Mibba	ZH	F	46	155	467
ZH Baseline	Lilly	ZH	F	27	86	360
Total calls				**562**	**2272**	**3096**

3.5.2 Recording procedures and processing

For the vocal recordings, we used a Condenser Microphone CM16/CMPA and an AviSoft UltraSoundGate 116H together with the RECORDER software from AviSoft in all conditions. The microphone was directed toward the focal individual, which changed after 10 minutes. Recording sessions lasted 20 min, so every individual of a pair was the focal individual once per session. The order of focal animals was counterbalanced. The caller was identified visually by the observer and calls were labelled digitally during the recording using the labelling function provided by AviSoft RECORDER.

Each recording was visually scanned and each selected call saved as a separate sound file using AviSoft SASLab pro. Each single call was then measured in Praat [Boersma and Weenink, 2009], using a script by Reby and McComb ([Reby and McComb, 2003], adapted by E. Briefer), and each call was again visually controlled to ensure correct measuring. We measured 15 (phee calls and food calls) or 17 (trill calls) parameters for each call. These were for each call the fundamental frequency (F0) at the start and end of the call, mean, minimal and maximal F0, percentage of the call duration for which F0 was max, the absolute slope of F0, mean variation of F0 per second, the frequency values at the first, second and third quartiles of energy, the highest frequency of the whole spectrum, percentage of time this highest

frequency is reached and jitter, as well as frequency modulation rate and frequency modulation extent for trill calls (for detailed description see [Briefer and McElligot, 2011; Zürcher and Burkart, 2017]).

For the analysis of experiment 1, we selected the first ten phee calls, five trills as well as twenty food calls from each recording of the ZH animals to reach a balanced sample. As the MA animals tended to vocalize only rarely in the "*NewPhys*" quarantine condition, we used every call they produced that passed our quality criteria for the subsequent analysis. Calls were excluded from the analysis if the identity of the caller could not be determined, if they overlapped with any other call or noise, if they were mixed with background noise or if we could not measure the whole call correctly in Praat. For the analysis of part 2, we only included animals in the data set if they contributed at least 5 calls per call type for a specific week.

3.5.3 Study setup

3.5.3.1 Part 1: Environmental accommodation

Environmental accommodation was tested in 12 animals in total in two different settings. Eight marmosets (ZH animals, six males, two females) were recorded before and after the whole colony was moved to a new building (translocation A in Figure 3.1), whereas the other four marmosets (MA animals) were recorded after moving from their natal colony in Madrid, Spain to a quarantine station in Switzerland before being introduced to the new colony (translocation B_1 in Figure 3.2).

The ZH animals were transferred from the old colony building ("*Before*" condition) to the new colony building ("*After1*" and "*After2*" condition). The new room was equipped with similar cages and enrichment as the old room, but differed in the surface materials of the walls and ceilings (metal and plastic vs wood) and in shape and size, supposedly resulting in rather different acoustic properties. To exclude social changes for the animals, the animals remained with their familiar partner in the new facilities and were housed in a similar social setting with their familiar neighbours.

All the animals were recorded repeatedly at three points in time: in the week before the moving process ("*Before*"), in the first week after the moving process ("*After1"*) and after five – six weeks ("*After2"*) (see Fig 1). In the "*Before*" condition, animals were shortly separated from their cage mate to elicit phee calls. In all other situations and for all other call types, animals were recorded in pairs. To elicit food calls, animals were presented with a mixture of highly preferred food such as meal worms, cashew nuts and gum in both conditions.

The MA animals were two males and two females, housed as a pair of brothers and a pair of sisters. They were initially living in their natal family group and were separated from their family either shortly before or at the time of the translocation B_1. They were recorded at three points in time, over two weeks in their natal colony and family, seven months before translocation ("*Before*" condition), over 4 weeks in the quarantine ("*NewPhys*" condition, after Translocation B_1) and over 16 weeks after introduction into the colony room of the new colony ("*NewSoc*", after Translocation B_2, see second part, social accommodation, and Fig 2). In the "*Before*" condition, the animals were recorded in pairs with another animal of their natal family group [see Zürcher and Burkart, 2017] and during the "*NewPhys"* and the "*NewSoc"* conditions with their sibling. To elicit food calls, animals were provided with different highly preferred food items like insects and pieces of banana. As the animals were rarely calling during the "*NewPhys*" situation, we

provided a short play-back stimulus twice (females) and three times (males) to elicit calling behaviour. As playback stimuli we chose phee calls and food calls from former colony members in MA.

3.5.3.2 Part 2: Social accommodation

Social accommodation was tested in the four MA animals. After the time in the quarantine, the MA animals were introduced into the colony room housing the ZH population (Translocation B_2). The animals were still housed with their respective sibling, and were visually and physically separated from the new population, but could vocally interact, therefore were exposed to a new vocal dialect [Zürcher and Burkart, 2017]. The four MA animals were recorded over a period of 16 weeks in their home cage ("*NewSoc")*, either alone to elicit phee calls or with their sibling. To elicit food calls, they were again provided with favourite food, like banana or insects.

The four MA animals were compared to a randomly chosen matched-control set of four ZH colony members. These animals, two males and two females, were unrelated, adult, non-breeding individuals that were all first housed with their family and later individually, as they were intended to become new breeding individuals (similar to the MA animals). Two of the ZH animals (one male and one female) were in acoustic contact with the MA animals, whereas the two other ZH animals were not, as they were housed in another colony room.

3.5.4 Statistics

For all the analyses, we first reduced the dimensionality of our data. Therefore, we performed a principal component analysis on z-transformed values for each of the measured parameters. We extracted components with Eigenvalues greater than the 95% quantile value obtained from 10000 randomly generated datasets with equal sample size and dimensionality as our empirical data.

All the analyses were performed with the statistics program "R". pDFA were performed after a script by R. Mundry [Mundry and Sommer, 2007]. For Eigenvalue extraction of the PCAs we used the package "nFactors"[Raiche, 2010], for calculating the LMEs the packages "lmerTest" [Kuznetsova et al., 2017] and "MuMIn" [Barton, 2009] and for creating the graphs "ggplot2" [Wickham, 2009].

3.5.4.1 Part 1: Environmental accommodation

To test whether calls were different before and after translocation between environments in the ZH animals, we performed a permutated Discriminant Function analysis (pDFA) [Mundry and Sommer, 2007]. We used *condition ("Before" and "After1")* as test factors and *individual* as control factor, and included the extracted PCA-factor as test variables. To test if changes were stable over time, we performed two crossed pDFAs, one for the ZH animals and one for the MA animals, with the condition "*Before*" and "*After*2" and "*Before*" and "*NewPhys*" as test factors and *individual* as control factors, including the extracted PCA-factor as test variables.

3.5.4.2 Part 2: Social accommodation

To assess whether the vocalisations of the four MA animals converged to their new population's average vocalisation, we first performed a PCA. Next we weighted these component by the proportion of the total variance they explained, and calculated the Euclidean distance between each recorded vocalisation and its corresponding average vocalisation of the Zurich population. Mathematically we defined these three

population average vocalisations as the centroids in the four-dimensional principal component space of each call type, and operationalised this as the average of the four Zurich individuals' 10% trimmed-means along each of the component axes. We thus ended up with a single continuous value to express the extent to which each vocalisation of the MA animals differed from its call type specific average of the Zurich population, *i.e.* vocal distance.

To establish whether these vocal distances changed over time, we fitted a series of Linear Mixed Effects models (LME). Model parameters were approximated using maximum likelihood estimation, while model performance was assessed by likelihood-ratio tests against a null model consisting of the intercept and random effects only. In addition, the proportion of the total variance accounted for by each model was assessed by both marginal and conditional R^2 values[Nakagawa and Schielzeth, 2013]. Following statistical convention, we only interpreted the highest order effect in which a predictor variable occurred whenever a significant interaction was present. To ensure parametric assumptions were met, we ln-transformed the outcome variable in all analyses.

Our model aimed to express vocal distance as a function of sex and week since introduction (where week 0 signifies recordings prior to auditory exposure to the local population ("*Before*") and 0.5 signifies recordings in the "*NewPhys" condition after the first but before the second translocation*), and additionally considered second-order interactions between the two variables to investigate possible sex specific patterns.

To test whether calls were significantly different immediately before the second translocation and after 5+ weeks in the new colony, we performed a linear model expressing vocal distance as a function of condition (last two weeks before second translocation, *"NewPhys"* vs. 5+ weeks after translocation) and sex, as well as the interaction of the two. We choose 5+ weeks as a time range of comparison as we could show in the first part of the manuscript that potential changes in vocalizations due to the translocation should have been lost by then.

We used different statistical methods in experiment 1 and 2 due to the different questions we aimed to answer with the respective experiment. In experiment 1, we compared calls of the same individuals between different situations, and not to a specific baseline. We therefore applied the method commonly used for this type of comparison. For the experiment 2, we were not only interested in whether the calls of the MA animals change, but whether they became more similar to each other. We therefore developed a method to measure vocal distance between several individuals, to test for changes therein.

3.5.5 Acknowledgements

We are thankful to A. Fidalgo and S. Sanches for the four monkeys we got from their colony and Zoo Zurich for providing temporary care and accommodation for the animals. We would like to thank all caretakers and research assistant that were involved with the procedures of collecting or processing of the data, as well as C. van Schaik for comments and suggestions. We would also like to thank the editor, Dr. Sakata, as well as the anonymous reviewers for their constructive comments about earlier stages of this manuscript.

3.6 Supplementary material

3.6.1 Experiment 1, Environmental accommodation

Table S3.5: **Factor loadings of the call parameters** on the extracted PCA factors (PC 1- PC 4) for the ZH animals in the analysis of the Experiment 1: Environmental accommodation.

	Trill calls			**Phee calls**				**Food calls**			
Parameter	PC1	PC2	PC3	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
sound duration	-0.26	-0.15	-0.10	-0.07	0.46	-0.13	0.08	0.04	0.52	-0.07	0.03
Mean F0	-0.39	0.19	0.07	-0.43	-0.02	0.08	-0.01	0.43	0.00	0.12	0.04
F0 start	-0.25	0.41	0.13	-0.37	0.07	0.15	0.31	0.42	-0.01	0.11	-0.11
F0 end	-0.41	0.08	0.05	-0.37	0.00	-0.06	-0.34	0.39	0.08	0.13	0.24
Max F0	-0.42	0.07	0.04	-0.41	-0.04	0.02	-0.15	0.42	-0.01	0.11	-0.10
percTime Max F0	-0.17	-0.15	0.10	-0.06	0.22	-0.34	-0.66	-0.04	-0.44	0.02	0.47
Min F0	-0.25	0.41	0.14	-0.39	0.11	0.12	0.18	0.39	0.08	0.13	0.24
F0 abs slope	-0.21	-0.35	-0.04	0.00	-0.53	0.11	-0.04	0.13	-0.44	0.06	-0.32
F0 Var	-0.25	-0.37	-0.07	-0.02	-0.52	0.11	-0.07	0.07	0.02	-0.02	-0.66
Q25perc	-0.06	-0.08	0.57	-0.40	-0.05	-0.07	0.18	0.31	-0.07	-0.34	0.03
Q50perc	0.14	-0.21	0.53	-0.16	-0.24	-0.54	0.11	0.15	-0.15	-0.62	0.00
Q75perc	0.04	-0.20	0.51	0.04	-0.25	-0.58	0.16	0.08	-0.07	-0.61	0.01
Fpeak	-0.06	-0.05	-0.01	0.00	0.12	-0.25	0.06	0.02	0.15	-0.01	0.04
percTime of max intensity	-0.15	-0.02	-0.01	0.01	0.09	-0.32	0.39	-0.05	-0.34	0.08	0.22
jitter	0.21	0.02	-0.14	-0.14	-0.16	-0.06	-0.26	0.02	-0.41	0.20	-0.22
Fm Rate	0.15	0.29	0.17								
FM extend	-0.24	-0.37	-0.13								
Eigenvalue	5.46	3.35	2.15	5.00	3.14	1.46	1.15	5.25	2.16	1.83	1.68
% of total variance	32.12	19.70	12.63	33.36	20.90	9.72	7.70	35.03	14.41	12.18	11.22

Table S3.6: **Factor loadings of the call parameters** on the extracted PCA factors (PC 1- PC 4) for the MA animals in the analysis of the Experiment 1: Environmental accommodation.

	Trill calls				**Phee calls**				**Food calls**			
Parameter	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
sound duration	0.08	-0.15	-0.32	-0.49	0.12	-0.45	0.07	-0.10	-0.06	-0.36	-0.07	0.22
Mean F0	-0.40	-0.06	0.01	-0.07	0.42	-0.06	0.10	0.08	-0.42	0.04	0.03	0.10
F0 start	-0.38	-0.04	0.12	-0.02	0.32	0.27	-0.20	0.25	-0.40	0.10	-0.11	0.12
F0 end	-0.37	-0.14	-0.03	-0.05	0.38	-0.18	0.19	-0.04	-0.40	-0.04	0.21	0.08
Max F0	-0.39	0.01	-0.01	-0.14	0.38	-0.16	0.22	0.05	-0.40	0.08	-0.13	0.15
percTime Max F0	0.02	-0.19	-0.12	0.17	0.03	-0.26	0.22	-0.49	0.02	-0.44	0.02	0.29
Min F0	-0.38	-0.15	0.09	0.05	0.35	0.22	-0.21	0.15	-0.40	-0.02	0.22	0.06
F0 abs slope	-0.08	0.53	-0.08	0.12	-0.07	0.31	0.57	0.03	0.01	0.06	-0.64	0.28
F0 Var	-0.09	0.52	0.04	0.01	-0.04	0.24	0.63	0.06	0.02	0.49	-0.42	-0.22
Q25perc	-0.39	-0.06	0.07	0.01	0.41	0.01	-0.01	0.06	-0.33	-0.05	-0.05	-0.35
Q50perc	-0.26	-0.02	-0.38	0.19	0.33	0.15	0.04	-0.29	-0.26	-0.27	-0.30	-0.41
Q75perc	-0.04	-0.10	-0.54	0.29	0.10	0.33	-0.19	-0.51	0.03	-0.47	-0.15	-0.53
Fpeak	0.02	-0.09	-0.20	-0.41	0.04	-0.26	-0.02	-0.15	-0.05	0.02	-0.09	-0.08
percTime of max intensity	-0.05	0.20	0.06	0.39	0.00	-0.05	-0.06	-0.42	-0.11	0.21	-0.05	-0.06
jitter	0.04	0.01	-0.53	0.26	0.04	0.43	-0.03	-0.32	0.00	-0.28	-0.40	0.33
Fm Rate	0.03	-0.19	0.26	0.34								
FM extend	-0.11	0.49	-0.14	-0.26								
Eigenvalue	6.00	2.60	2.07	1.60	5.28	3.21	1.80	1.21	5.59	2.07	1.69	1.24
% of total variance	35.32	15.30	12.19	9.44	35.19	21.42	11.99	8.06	37.30	13.82	11.24	8.29

3.6.2 Experiment 2, Social accommodation

Table S3.7: Factor loadings of the call parameters on the extracted PCA factors (PC 1- PC 4) for the MA and ZH baseline animals used in the analysis of the Experiment 2: Social accommodation.

	Trill calls				**Phee calls**				**Food calls**			
	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
sound duration	0.09	-0.27	-0.40	0.30	0.04	-0.53	0.11	-0.01	0.07	-0.08	-0.39	0.33
Mean F0	-0.43	-0.16	0.00	0.03	0.40	-0.05	0.11	-0.15	0.41	0.09	0.07	0.03
F0 start	-0.43	0.09	0.14	-0.01	0.35	0.16	-0.24	-0.18	0.41	-0.06	0.13	0.06
F0 end	-0.37	-0.26	-0.15	0.09	0.36	-0.16	0.17	-0.07	0.36	0.33	0.00	0.02
Max F0	-0.38	-0.29	-0.03	0.04	0.36	-0.15	0.27	-0.13	0.41	-0.05	0.12	0.04
percTime Max F0	0.10	-0.21	-0.35	0.11	0.02	-0.33	0.20	0.27	-0.03	0.25	-0.19	-0.59
Min F0	-0.44	0.10	0.07	0.02	0.36	0.13	-0.25	-0.15	0.36	0.32	0.00	0.03
F0 abs slope	0.09	-0.28	0.45	-0.29	-0.05	0.42	0.50	-0.04	0.14	-0.44	0.31	-0.21
F0 Var	0.10	-0.37	0.31	-0.28	-0.04	0.37	0.55	-0.07	0.07	-0.46	0.34	0.19
Q25perc	-0.21	-0.12	0.31	0.30	0.39	-0.02	0.06	0.00	0.37	-0.09	-0.17	-0.06
Q50perc	0.18	-0.04	0.37	0.45	0.35	0.09	0.10	0.27	0.25	-0.34	-0.34	-0.27
Q75perc	0.11	0.06	0.28	0.54	0.15	0.23	-0.14	0.59	0.04	-0.36	-0.49	-0.27
Fpeak	0.04	-0.06	-0.10	0.36	0.02	-0.19	0.09	0.23	0.03	-0.10	-0.11	0.00
percTime of max intensity	0.04	-0.19	0.02	0.02	0.04	-0.06	0.12	0.57	0.02	0.17	0.26	-0.44
jitter	-0.06	0.24	0.20	0.09	0.12	0.32	-0.31	0.16	-0.04	-0.12	0.32	-0.35
Fm Rate	-0.07	0.36	0.00	-0.06								
FM extend	0.11	-0.47	0.05	-0.01								
Eigenvalue	4.70	3.2	1.96	1.45	5.58	2.61	1.92	1.16	5.61	2.05	1.57	1.3
% of total variance	27.7	18.9	11.5	8.5	37.2	17.4	12.8	7.8	37.4	13.7	10.5	9.3

Table S3.8: **Social accommodation after 5+ weeks:** The effect of condition (shortly before translocation or 5+ weeks after translocation), sex, and the respective interactions on vocal distance (ln-transformed Euclidean distance) between the vocalisations of four translocated individuals and the call type specific average of the new population. Parameter estimates, standard errors, and statistical significance are obtained from a linear mixed effects model. Significant (highest-order) effects are indicated with p-values in bold.

Trill

	B	**SE**	**t**	**P**
Intercept	0.5604	0.2536		
Condition	-0.9590	0.2459	-3.9	< 0.001
Sex	-0.6063	0.2801	-2.164	0.046
Condition * Sex				
Before and after translocation by sex	0.7083	0.2530	2.8	**0.006**

$N_{obs.}$= 515 from 4 individuals; R^2_m= 0.128, R^2_c= 0.251 χ^2_{MLT}= 8.095, p< 0.05

Phee

	B	**SE**	**t**	**P**
Intercept	0.05282	0.09502		
Condition	-0.34725	0.05295	-6.558	< 0.001
Sex	-0.44615	0.13333	-3.346	0.048
Condition * Sex				
Before and after translocation by sex	0.27506	0.07434	3.700	**< 0.001**

$N_{obs.}$= 1001 from 4 individuals; R^2_m= 0.085, R^2_c= 0.135, χ^2_{MLT}= 12.7693 , p< 0.05

Food

	B	**SE**	**t**	**P**
Intercept	-0.709	0.084		
Condition	0.247	0.069	3.561	**< 0.001**

$N_{obs.}$= 591 from 3 individuals; R^2_m= 0.0244, R^2_c= 0.0504, χ^2_{MLT}= 9.846, p< 0.05

3.6.3 Parameter changes due to translocations

We tested changes in vocal distance in call structure based on the combination of all call parameters, therefore it is not straightforward to identify which parameter contributed most to the change in vocal distance. We nevertheless provide the average parameter for each individual and each call type before and after the translocation in the supplementary material 2 (Table S2.1 – S2-4) for an overview, but would like to emphasize that it is the combination of all rather than a single parameter that was responsible for the observed results. Changes in any single parameter therefore only have a limited informative value.

Table S3.9: Average value of each call parameter from the ZH animal in experiment 1 from "Before" and "After1"

		Kapi		Kapo		Tabor		Thilo		Kaliper		Marlene		Vesta		Vito	
		Before	*After1*	*Before*	*After1*	*Before*	*After1*	*Before*	*After1*	*Before*	*After1*	*Before*	*After1*	*Before*	*After1*	*Before*	*After1*
Trill calls	sound duration	0.472	0.522	0.446	0.599	0.583	0.565	0.498	0.508	0.517	0.579	0.645	0.735	0.415	0.383	0.776	0.451
	Mean F0	6733.6	5516.2	6036.9	6068.1	6376.5	6582.8	6631.9	6634.3	6826.3	6450.0	6561.4	6353.6	5979.2	5909.6	6923.8	6055.9
	F0 start	6198.9	4890.4	5610.8	5357.1	5355.1	5724.6	6019.5	6342.7	6536.9	5916.2	5850.3	5432.8	5632.0	5679.6	5168.3	5507.8
	F0 end	7025.2	5718.4	6086.5	6333.3	6961.5	6925.2	7057.1	6916.0	6877.7	6602.8	6982.8	6784.7	6249.9	6079.5	7712.5	6251.2
	Max F0	7120.7	5908.6	6255.0	6417.8	7101.7	7074.8	7122.8	6939.7	7077.3	6826.6	7037.2	6902.8	6303.6	6186.3	7839.0	6321.8
	% Time Max F0	77.1	72.9	66.7	76.3	73.9	82.1	81.4	66.1	67.4	69.3	77.7	80.3	68.2	60.1	83.0	71.3
	Min F0	6180.1	4869.7	5592.0	5357.1	5292.8	5724.6	6015.0	6231.9	6493.0	5833.1	5842.4	5406.9	5597.5	5613.6	5168.3	5503.7
	F0 abs slope	2958.0	3073.0	2385.0	2654.5	3957.0	3424.8	3105.2	2392.2	2251.0	2864.5	2470.2	2761.0	2165.7	2301.6	4308.5	2762.8
	F0 Var	1883.1	2150.6	1405.3	1874.3	2821.7	2356.0	2126.8	1431.2	1417.1	1921.9	1656.0	2049.2	1242.9	1318.0	3348.4	1688.7
	Q25%	8017.7	7586.3	6446.2	6746.6	6876.4	7959.0	6488.2	6658.3	7093.1	6747.6	6702.5	7637.5	6290.2	8454.0	7035.5	6623.3
	Q50%	10822.4	11807.5	9977.1	10265.0	9265.7	9630.5	7861.4	8216.6	8756.7	8888.5	9293.8	9123.6	9563.2	11241.1	8137.6	9696.6
	Q75%	15473.7	16793.8	12508.1	13324.6	12657.5	12845.7	11361.9	13216.4	12689.6	13287.1	12841.3	12085.5	12978.0	14749.5	11919.1	12492.1
	Fpeak	200.6	229.9	214.2	204.7	219.5	222.4	174.0	158.4	229.7	215.7	203.1	211.1	203.2	198.8	338.2	217.5
	% Time of max intensity	22.8	28.8	28.9	53.0	35.2	19.5	26.1	28.9	29.7	32.1	33.5	49.8	28.0	25.2	65.9	26.9
	jitter	0.019	0.021	0.023	0.027	0.019	0.023	0.023	0.016	0.021	0.018	0.024	0.020	0.023	0.023	0.015	0.020
	Fm Rate	2.2	1.7	2.0	2.0	1.5	2.6	2.1	2.2	2.9	2.0	1.8	1.2	2.5	2.7	0.8	2.0
	FM extend	1355.9	1664.3	1117.8	1223.8	2822.0	1405.4	1432.3	727.4	744.3	1271.9	1468.3	2287.1	949.6	601.9	4724.1	994.0
Phee Calls	sound duration	1.089	0.971	0.962	1.165	0.787	0.878	1.231	1.014	1.305	1.358	1.097	1.324	1.963	1.318	1.137	1.672
	Mean F0	7244.2	6804.8	7747.3	7359.7	6979.1	6977.2	7507.5	7560.4	7856.1	7455.8	7690.0	7546.2	7095.3	7214.6	7957.3	7765.0
	F0 start	6518.2	5866.6	6833.5	6276.5	6238.7	5983.4	6623.9	6553.1	7130.3	6634.4	6855.1	6736.4	6511.8	6234.8	7310.6	6765.2
	F0 end	7627.5	7338.1	8119.8	7806.6	6899.9	7003.9	7592.4	7824.8	8168.0	7742.8	8033.1	7701.9	6719.9	7319.5	8165.3	7862.4
	Max F0	7858.1	7477.6	8393.8	8058.1	7450.3	7438.7	8116.8	8118.4	8438.6	8061.4	8272.5	8127.7	7430.9	7663.0	8479.3	8116.0
	% Time Max F0	84.0	85.2	79.4	82.9	71.6	77.3	79.4	79.8	86.6	84.7	77.2	78.3	86.6	83.8	80.5	71.3
	Min F0	6478.8	5859.4	6754.1	6251.2	6190.6	5979.6	6545.7	6549.7	7127.5	6629.2	6791.7	6731.9	6359.1	6229.4	7115.0	6735.8

	F0 abs slope	2146.6	2725.5	2892.4	2719.9	3022.1	2895.4	2200.6	2524.8	1454.3	1505.2	2293.4	1758.2	982.2	1767.2	1827.9	1381.9
	F0 Var	1691.5	2171.6	2209.9	2123.2	2202.5	2167.5	1734.4	1889.0	1169.3	1212.6	1844.2	1347.1	687.2	1391.6	1483.2	1091.0
	Q25%	7195.4	6816.4	7700.6	7351.8	6976.3	7004.5	7314.8	7408.4	7690.7	7370.8	7479.8	7553.1	7126.9	7287.9	7729.7	7753.8
	Q50%	8151.8	8358.6	8596.7	8472.0	7557.0	7705.0	7713.7	7794.9	7995.7	7940.8	7856.9	8679.3	7333.5	8097.6	8192.1	7943.9
	Q75%	13085.6	13276.4	12133.5	12678.7	10402.2	11439.4	9525.2	9399.7	9097.2	10925.2	8482.6	13598.8	10371.8	12785.2	10460.8	9029.7
	Fpeak	234.9	277.5	233.9	263.9	212.0	231.6	255.2	252.8	257.1	346.4	229.7	286.0	276.7	289.1	253.4	245.3
	% Time of max intensity	46.8	48.0	48.5	47.9	48.0	48.3	39.5	41.1	46.4	46.6	47.6	50.1	52.1	58.8	32.3	53.9
	jitter	0.007	0.005	0.008	0.007	0.006	0.005	0.006	0.007	0.006	0.006	0.006	0.005	0.005	0.005	0.012	0.007
Food Call	sound duration	0.099	0.107	0.144	0.152	0.114	0.106	0.101	0.101	0.117	0.120	0.238	0.229	0.121	0.120	0.112	0.115
	Mean F0	7448.5	6527.5	8138.7	8121.5	7301.1	6584.5	7841.2	7382.4	7984.1	7480.4	7942.5	6553.5	8348.6	7999.4	7031.2	6822.1
	F0 start	8067.1	7218.1	8895.4	8867.7	7946.7	7185.6	8477.5	8123.7	8663.7	8205.0	8739.1	7440.5	9047.2	8789.7	7628.1	7398.0
	F0 end	6819.3	5821.7	7067.7	7203.0	6584.6	5843.2	7138.2	6575.1	7314.0	6736.3	7302.9	5894.2	7586.1	7122.9	6417.0	6225.8
	Max F0	8067.1	7219.0	8903.1	8881.2	7959.9	7200.0	8486.0	8124.8	8664.2	8205.0	8751.4	7446.5	9048.2	8790.1	7628.1	7399.3
	% Time Max F0	27.8	28.0	22.8	20.3	25.7	26.2	27.7	29.0	24.0	24.3	18.3	15.5	24.4	24.1	26.3	24.8
	Min F0	6819.3	5821.7	7067.7	7202.4	6580.2	5836.8	7138.2	6574.5	7314.0	6736.2	7297.4	5890.0	7586.1	7122.9	6417.0	6225.8
	F0 abs slope	34734.7	37231.5	33278.0	28902.9	30543.6	28568.0	43676.9	43796.5	27748.5	29509.3	14781.1	13924.2	33416.2	36044.3	31096.8	25747.0
	F0 Var	3816.9	3777.7	6666.6	6844.7	4866.8	5948.4	3726.6	5064.1	5485.9	6172.5	4451.8	6036.7	4856.5	5109.4	3251.4	4042.3
	Q25%	7554.2	6809.7	7887.3	7951.6	7525.0	7731.8	7539.9	7239.5	7636.1	7397.6	7831.4	6836.6	7630.9	7783.1	6872.5	6817.1
	Q50%	9057.1	9160.1	9577.3	9498.3	9773.9	10402.1	9284.7	9100.3	8600.0	8912.2	8919.7	9415.2	9014.5	9292.2	8092.1	8450.8
	Q75%	12558.7	14194.1	13684.2	13964.8	13884.2	13675.0	13272.2	12366.2	11502.0	13466.7	12746.3	13439.9	11630.8	13982.6	11227.0	11862.0
	Fpeak	170.6	175.2	186.6	177.4	164.6	160.8	145.8	187.6	159.3	177.1	197.7	199.9	164.1	187.8	175.5	174.6
	% Time of max intensity	41.6	44.3	44.7	41.5	46.0	48.9	44.0	44.4	41.8	43.8	36.9	36.5	48.7	44.3	43.7	45.1
	jitter	0.032	0.028	0.025	0.024	0.025	0.023	0.027	0.024	0.024	0.026	0.015	0.015	0.023	0.020	0.028	0.024

Table S3.10: Average value of each call parameter from the ZH animals in experiment 1 from "Before" and "After2"

		Kapi		Kapo		Tabor		Thilo		Kaliper		Marlene		Vesta		Vito	
		Before	*After2*	*Before*	*After2*	*Before*	*After2*	*Before*	*After2*	*Before*	*After2*	*Before*	*After2*	*Before*	*After2*	*Before*	*After2*
Trill calls	sound duration	0.472	0.477	0.446	0.387	0.583	0.570	0.498	0.468	0.517	0.501	0.645	0.596	0.415	0.344	0.776	0.360
	Mean F0	6733.6	6149.8	6036.9	5444.6	6376.5	5864.6	6631.9	6542.2	6826.3	6439.8	6561.4	6161.5	5979.2	6013.9	6923.8	6101.2
	F0 start	6198.9	5670.1	5610.8	5196.0	5355.1	4809.7	6019.5	5906.2	6536.9	6177.0	5850.3	5502.3	5632.0	5904.1	5168.3	5736.5
	F0 end	7025.2	6270.3	6086.5	5384.6	6961.5	6417.5	7057.1	6941.5	6877.7	6456.3	6982.8	6472.6	6249.9	6145.6	7712.5	6192.5
	Max F0	7120.7	6437.5	6255.0	5587.1	7101.7	6459.4	7122.8	6978.1	7077.3	6630.1	7037.2	6566.1	6303.6	6221.9	7839.0	6284.5
	% Time Max F0	77.1	67.1	66.7	57.2	73.9	78.6	81.4	79.7	67.4	54.5	77.7	78.2	68.2	64.0	83.0	68.9
	Min F0	6180.1	5630.2	5592.0	5148.9	5292.8	4809.7	6015.0	5904.8	6493.0	6112.3	5842.4	5493.7	5597.5	5847.3	5168.3	5732.3
	F0 abs slope	2958.0	2651.3	2385.0	2190.1	3957.0	3550.2	3105.2	3150.6	2251.0	1904.0	2470.2	2387.1	2165.7	1943.4	4308.5	2322.7
	F0 Var	1883.1	1585.7	1405.3	1187.6	2821.7	2669.2	2126.8	2081.4	1417.1	1191.4	1656.0	1643.9	1242.9	895.0	3348.4	1365.1
	Q25%	8017.7	7940.5	6446.2	6564.9	6876.4	6642.8	6488.2	6471.7	7093.1	6454.5	6702.5	7159.8	6290.2	8032.8	7035.5	6885.2
	Q50%	10822.4	11494.8	9977.1	9669.0	9265.7	9606.7	7861.4	8229.4	8756.7	8679.4	9293.8	9987.3	9563.2	11160.6	8137.6	9966.6
	Q75%	15473.7	14959.2	12508.1	12075.6	12657.5	12252.4	11361.9	11779.1	12689.6	11935.7	12841.3	12271.3	12978.0	14502.4	11919.1	13274.6
	Fpeak	200.6	219.5	214.2	174.8	219.5	205.0	174.0	266.7	229.7	200.1	203.1	224.4	203.2	201.7	338.2	180.6
	% Time of max intensity	22.8	24.4	28.9	30.9	35.2	25.7	26.1	22.2	29.7	29.3	33.5	33.1	28.0	27.5	65.9	24.9
	jitter	0.019	0.022	0.023	0.028	0.019	0.025	0.023	0.019	0.021	0.024	0.024	0.022	0.023	0.024	0.015	0.024
	Fm Rate	2.2	2.4	2.0	2.3	1.5	1.8	2.1	1.7	2.9	2.5	1.8	1.6	2.5	2.8	0.8	2.2
	FM extend	1355.9	1113.7	1117.8	728.5	2822.0	2196.4	1432.3	1621.2	744.3	613.8	1468.3	1579.8	949.6	444.9	4724.1	1030.8
Phee Calls	sound duration	1.089	1.040	0.962	1.172	0.787	1.044	1.231	0.949	1.305	1.374	1.097	1.469	1.963	1.442	1.137	1.546
	Mean F0	7244.2	6809.9	7747.3	7399.6	6979.1	6789.1	7507.5	7506.5	7856.1	6950.7	7690.0	7767.3	7095.3	7409.1	7957.3	8053.2
	F0 start	6518.2	6138.4	6833.5	6381.8	6238.7	5952.1	6623.9	6491.3	7130.3	6368.5	6855.1	7179.0	6511.8	6897.1	7310.6	7055.6
	F0 end	7627.5	7504.7	8119.8	7789.5	6899.9	6885.4	7592.4	8086.8	8168.0	7240.1	8033.1	7906.7	6719.9	7798.6	8165.3	8046.6
	Max F0	7858.1	7590.6	8393.8	8053.4	7450.3	7241.8	8116.8	8231.6	8438.6	7342.0	8272.5	8240.5	7430.9	8091.5	8479.3	8591.8
	% Time Max F0	84.0	85.0	79.4	81.4	71.6	80.2	79.4	83.7	86.6	87.2	77.2	76.8	86.6	76.1	80.5	87.4
	Min F0	6478.8	6101.8	6754.1	6381.1	6190.6	5931.6	6545.7	6491.3	7127.5	6368.5	6791.7	7075.8	6359.1	6608.4	7115.0	6984.6
	F0 abs slope	2146.6	2132.1	2892.4	2327.1	3022.1	2180.8	2200.6	2529.9	1454.3	993.7	2293.4	1299.3	982.2	1683.2	1827.9	1751.9

	F0 Var	1691.5	1606.4	2209.9	1769.0	2202.5	1601.2	1734.4	2003.3	1169.3	792.0	1844.2	1083.1	687.2	1375.0	1483.2	1314.2
	Q25%	7195.4	6710.8	7700.6	7339.8	6976.3	7003.5	7314.8	7265.5	7690.7	6921.2	7479.8	7731.5	7126.9	7301.1	7729.7	7937.6
	Q50%	8151.8	9044.4	8596.7	8157.2	7557.0	7904.7	7713.7	7854.3	7995.7	8650.4	7856.9	8267.6	7333.5	7700.8	8192.1	8258.5
	Q75%	13085.6	13160.3	12133.5	12002.0	10402.2	12576.1	9525.2	10367.2	9097.2	12579.2	8482.6	12761.4	10371.8	11399.9	10460.8	9501.8
	Fpeak	234.9	306.3	233.9	256.5	212.0	258.5	255.2	242.3	257.1	215.0	229.7	317.4	276.7	268.4	253.4	235.7
	% Time of max intensity	46.8	44.0	48.5	48.9	48.0	58.6	39.5	39.8	46.4	45.7	47.6	46.9	52.1	48.4	32.3	49.5
	jitter	0.007	0.005	0.008	0.006	0.006	0.006	0.006	0.009	0.006	0.005	0.006	0.005	0.005	0.008	0.012	0.007
Food Call	sound duration	0.099	0.106	0.144	0.150	0.114	0.101	0.101	0.101	0.117	0.136	0.238	0.195	0.121	0.114	0.112	0.107
	Mean F0	7448.5	7100.7	8138.7	8041.1	7301.1	6748.1	7841.2	7743.0	7984.1	8557.2	7942.5	7407.6	8348.6	7826.1	7031.2	7009.6
	F0 start	8067.1	7708.7	8895.4	8907.8	7946.7	7423.4	8477.5	8548.4	8663.7	9430.6	8739.1	8019.8	9047.2	8436.2	7628.1	7601.2
	F0 end	6819.3	6442.6	7067.7	7001.9	6584.6	5895.4	7138.2	6860.2	7314.0	7547.3	7302.9	6930.1	7586.1	7135.7	6417.0	6365.3
	Max F0	8067.1	7708.7	8903.1	8916.6	7959.9	7436.1	8486.0	8549.3	8664.2	9434.1	8751.4	8031.1	9048.2	8447.0	7628.1	7622.7
	% Time Max F0	27.8	27.7	22.8	25.4	25.7	23.7	27.7	28.7	24.0	24.8	18.3	19.4	24.4	24.5	26.3	25.2
	Min F0	6819.3	6442.6	7067.7	7001.9	6580.2	5895.4	7138.2	6859.6	7314.0	7547.3	7297.4	6929.1	7586.1	7135.7	6417.0	6345.6
	F0 abs slope	34734.7	37533.0	33278.0	37110.2	30543.6	30958.6	43676.9	47140.1	27748.5	34204.0	14781.1	12508.3	33416.2	27700.1	31096.8	27553.7
	F0 Var	3816.9	3158.2	6666.6	4575.7	4866.8	8319.9	3726.6	4628.1	5485.9	6633.9	4451.8	4783.3	4856.5	5133.3	3251.4	5280.4
	Q25%	7554.2	7212.9	7887.3	8186.2	7525.0	7847.3	7539.9	7524.2	7636.1	8376.6	7831.4	7228.6	7630.9	7211.7	6872.5	6959.1
	Q50%	9057.1	9619.1	9577.3	10062.3	9773.9	10495.8	9284.7	9022.0	8600.0	9601.0	8919.7	7992.2	9014.5	8607.5	8092.1	8499.8
	Q75%	12558.7	13933.2	13684.2	15714.6	13884.2	14320.3	13272.2	12336.6	11502.0	12165.4	12746.3	11253.0	11630.8	11347.2	11227.0	12851.8
	Fpeak	170.6	165.3	186.6	172.8	164.6	165.0	145.8	176.5	159.3	160.0	197.7	186.3	164.1	153.1	175.5	191.7
	% Time of max intensity	41.6	45.8	44.7	40.2	46.0	48.5	44.0	44.2	41.8	38.7	36.9	39.2	48.7	51.8	43.7	47.5
	jitter	0.032	0.027	0.025	0.022	0.025	0.020	0.027	0.023	0.024	0.025	0.015	0.017	0.023	0.035	0.028	0.022

Table S3.11: Average value of each call parameter from the MA animals in experiment 1 from "Before" and "NewPhys".

		Conan		Craken		Washington		Wisconsin	
		Before	*NewPhys*	*Before*	*NewPhys*	*Before*	*NewPhys*	*Before*	*NewPhys*
Trill calls	sound duration	0.274	0.334	0.398	0.647	0.272	0.419	0.214	0.364
	Mean F0	7291.6	8362.9	7684.3	8556.6	7321.6	7063.3	7846.9	7414.4
	F0 start	7337.4	8259.7	7784.6	7883.0	7448.7	6921.6	7905.5	7203.8
	F0 end	7090.6	8427.4	7499.4	8554.0	7218.5	7075.5	7708.4	7430.1
	Max F0	7433.1	8484.5	7944.6	8919.5	7542.4	7176.6	8001.1	7530.7
	% Time Max F0	25.5	70.1	30.5	72.2	25.6	54.0	33.1	52.2
	Min F0	7090.6	8208.9	7381.6	7738.7	7142.6	6894.5	7668.5	7180.8
	F0 abs slope	2511.1	1642.8	2806.3	3257.0	2194.3	1780.9	2585.2	2283.8
	F0 Var	1441.6	945.6	1640.0	2104.5	1568.5	849.2	1664.6	1111.6
	Q25%	7339.3	8279.3	7664.0	8116.9	7296.6	6932.6	7770.0	7230.1
	Q50%	7785.3	9206.0	8149.4	9190.3	7681.2	7663.7	8244.9	7938.1
	Q75%	13205.8	18796.1	12568.3	16116.0	9051.9	13086.0	9795.7	13019.9
	Fpeak	213.0	227.9	206.5	297.0	193.3	202.9	169.9	198.5
	% Time of max intensity	40.1	24.3	28.1	45.8	25.7	22.5	33.1	23.8
	jitter	0.030	0.040	0.034	0.043	0.025	0.043	0.026	0.043
	Fm Rate	2.6	3.7	3.0	3.4	3.8	3.4	3.5	2.7
	FM extend	580.5	406.1	652.2	1094.8	470.8	287.9	638.4	519.7
Phee calls	sound duration	0.422	0.703	0.447	0.987	0.539	0.800	0.532	0.839
	Mean F0	7697.6	8111.2	7680.6	8313.3	7161.8	7417.3	7528.9	7692.3
	F0 start	7557.4	7571.8	7502.7	7554.5	6936.8	6886.9	7309.4	7197.1
	F0 end	7667.3	8338.8	7635.9	8563.0	7138.1	7704.8	7590.5	7996.0
	Max F0	7923.2	8412.0	7973.5	8749.5	7415.9	7803.0	7775.5	8082.4
	% Time Max F0	54.7	75.0	37.5	72.5	60.9	79.9	66.9	79.6
	Min F0	7358.9	7548.5	7248.5	7514.8	6813.9	6866.3	7231.7	7181.2
	F0 abs slope	2319.4	1680.3	2720.1	1966.7	2073.9	2315.3	1789.8	2098.0
	F0 Var	1726.2	1337.3	1836.2	1489.8	1707.9	1408.8	1476.2	1447.5
	Q25%	7699.6	7932.2	7560.5	8001.7	7044.7	7174.2	7396.2	7503.3
	Q50%	8080.9	8290.3	8373.5	8585.8	7263.1	7455.5	7633.5	7849.5
	Q75%	12431.5	10284.2	12309.1	11570.2	8183.9	8530.1	9562.5	10562.4
	Fpeak	205.5	219.7	225.3	242.3	210.1	239.1	219.1	227.3
	% Time of max intensity	35.5	42.9	37.5	33.4	41.6	27.6	44.5	36.9
	jitter	0.029	0.016	0.028	0.025	0.011	0.019	0.019	0.021
Food calls	sound duration	0.064	0.090	0.078	0.093			0.084	0.101
	Mean F0	7892.3	8259.3	7543.4	7263.6			7623.8	7994.9
	F0 start	8239.8	8574.1	7905.2	7639.9			8037.4	8388.4
	F0 end	7471.5	7914.4	7122.8	6860.8			7211.5	7592.3
	Max F0	8281.6	8576.9	7975.7	7648.4			8050.8	8388.9
	% Time Max F0	21.7	27.1	23.8	24.6			20.0	29.6

Min F0	7435.6	7914.4	7118.1	6860.8	7206.4	7592.3
F0 abs slope	25835.3	23996.9	19278.0	31121.8	17044.9	27056.9
F0 Var	8636.1	1595.3	8047.0	1183.5	7333.6	1263.7
Q25%	7406.3	7356.7	7332.2	7349.6	7294.0	7290.3
Q50%	8221.8	8209.2	7953.1	8882.0	8100.4	8294.2
Q75%	9762.2	9537.9	10549.3	14215.8	11086.8	11320.9
Fpeak	161.9	165.9	146.0	163.0	153.0	163.9
% Time of max intensity	50.6	47.6	47.0	43.6	48.0	44.7
jitter	0.030	0.049	0.035	0.053	0.030	0.041

Table S3.12: Average value of each call parameter for each MA individual in the situation "Before" and "NewSoc" as well as of the ZH animals (Baseline) involved in experiment 2.

		Conan		Craken		Washington		Wisconsin		Lilly	Gatto	Mibba	Nautilus
		Before	*NewSoc*	*Before*	*NewSoc*	*Before*	*NewSoc*	*Before*	*NewSoc*	*Baseline*	*Baseline*	*Baseline*	*Baseline*
Trill calls	sound duration	0.274	0.634	0.398	0.738	0.272	0.516	0.214	0.502	0.570	0.369	0.702	0.808
	Mean F0	7291.6	7254.5	7684.3	7273.1	7321.6	7010.8	7846.9	7355.6	6609.6	6684.0	6652.1	7042.3
	F0 start	7337.4	6704.6	7784.6	6399.2	7448.7	6607.6	7905.5	6901.5	6191.9	6439.6	5956.3	6320.6
	F0 end	7090.6	7296.5	7499.4	7567.4	7218.5	7120.6	7708.4	7439.2	6688.2	6769.2	6812.9	7393.1
	Max F0	7433.1	7564.9	7944.6	7697.1	7542.4	7217.1	8001.1	7553.6	6957.9	6836.4	7090.9	7555.6
	Perc Time Max F0	25.5	63.0	30.5	74.9	25.6	69.2	33.1	65.3	62.2	64.3	65.9	74.8
	Min F0	7090.6	6647.6	7381.6	6393.9	7142.6	6603.7	7668.5	6886.9	6086.9	6432.9	5826.7	6284.5
	F0 abs slope	2511.1	2639.1	2806.3	2394.3	2194.3	1989.7	2585.2	2342.2	2831.2	2052.6	2629.1	2471.9
	F0 Var	1441.6	1625.9	1640.0	1776.7	1568.5	1212.4	1664.6	1395.5	1688.8	1114.4	1961.8	1666.5
	Q25perc	7339.3	7077.8	7664.0	6977.1	7296.6	6821.8	7770.0	7093.7	8148.4	6731.1	6787.9	7123.6
	Q50perc	7785.3	7797.6	8149.4	7583.1	7681.2	7455.9	8244.9	7579.8	9878.8	8633.1	9234.4	8072.6
	Q75perc	13205.8	10030.5	12568.3	9888.1	9051.9	10284.9	9795.7	9482.6	12738.7	13279.8	13103.2	12240.7
	Fpeak	213.0	201.6	206.5	191.4	193.3	186.1	169.9	186.9	235.9	210.3	236.1	220.9
	Perc Time of max intensity	40.1	26.3	28.1	30.0	25.7	23.1	33.1	22.2	33.3	26.9	24.7	54.4
	jitter	0.030	0.030	0.034	0.027	0.025	0.040	0.026	0.037	0.026	0.023	0.016	0.022
	Fm Rate	2.6	2.8	3.0	2.6	3.8	2.9	3.5	2.7	2.8	2.9	1.7	1.9
	FM extend	580.5	1026.5	652.2	1119.0	470.8	610.3	638.4	721.5	1244.1	525.3	1843.6	1855.9
Phee calls	sound duration	0.422	0.725	0.447	1.378	0.539	1.502	0.532	1.182	0.863	0.757	0.955	0.927
	Mean F0	7697.6	8006.5	7680.6	7660.0	7161.8	7464.9	7528.9	7344.7	7141.2	7296.1	7692.7	7512.6
	F0 start	7557.4	7422.0	7502.7	6864.4	6936.8	6585.0	7309.4	6627.7	6348.5	6609.7	6699.9	6625.7
	F0 end	7667.3	8165.3	7635.9	7777.7	7138.1	7859.9	7590.5	7619.1	7614.9	7226.1	8047.1	7793.5
	Max F0	7923.2	8300.8	7973.5	8065.5	7415.9	8024.8	7775.5	7872.2	7715.2	7656.6	8358.7	8224.8

	percTime Max F0	54.7	74.8	37.5	74.9	60.9	85.1	66.9	78.9	81.5	64.5	82.0	79.8
	Min F0	7358.9	7391.8	7248.5	6817.4	6813.9	6571.9	7231.7	6604.9	6316.7	6572.9	6673.8	6576.4
	F0 abs slope	2319.4	1847.3	2720.1	1612.6	2073.9	1422.6	1789.8	1943.4	2240.0	2580.2	2503.8	3049.4
	F0 Var	1726.2	1432.9	1836.2	1284.5	1707.9	1161.3	1476.2	1543.2	1807.5	2018.8	2012.2	2204.6
	Q25perc	7699.6	7826.0	7560.5	7471.7	7044.7	7218.2	7396.2	7094.9	7187.4	7176.9	7582.9	7363.4
	Q50perc	8080.9	8179.0	8373.5	7785.7	7263.1	7465.5	7633.5	7386.1	8079.7	7436.9	8285.8	7725.3
	Q75perc	12431.5	9817.2	12309.1	9009.0	8183.9	7906.2	9562.5	8298.8	12734.9	7912.3	11777.0	9214.9
	Fpeak	205.5	188.1	225.3	214.6	210.1	218.4	219.1	215.4	224.7	218.5	231.1	253.2
	percTime of max intensity	35.5	43.9	37.5	49.5	41.6	40.7	44.5	40.2	58.9	46.8	53.4	51.5
	jitter	0.029	0.018	0.028	0.014	0.011	0.011	0.019	0.012	0.004	0.005	0.007	0.007
	sound duration	0.064	0.115	0.078	0.132	0.084	0.141	0.084	0.143	0.156	0.105	0.122	0.135
	Mean F0	7892.3	8484.7	7543.4	7783.6	7830.2	9300.1	7623.8	7860.0	7947.8	7549.6	7955.0	6785.4
	F0 start	8239.8	8894.7	7905.2	8221.4	8157.1	9973.1	8037.4	8375.3	8791.3	7989.7	8519.9	7603.9
	F0 end	7471.5	7814.5	7122.8	7144.2	7502.7	8355.4	7211.5	7274.8	6770.2	7045.9	7227.1	5798.9
	Max F0	8281.6	8969.2	7975.7	8287.1	8164.7	9990.9	8050.8	8396.3	8853.4	8002.4	8542.4	7616.5
	percTime Max F0	21.7	28.3	23.8	25.8	19.8	23.8	20.0	22.0	21.3	24.4	21.4	23.3
	Min F0	7435.6	7798.4	7118.1	7133.4	7499.6	8355.4	7206.4	7264.7	6764.4	7041.9	7222.3	5798.9
Food Calls	F0 abs slope	25835.3	26740.0	19278.0	22176.7	11980.1	28653.7	17044.9	20804.3	30917.6	26234.2	25412.5	36039.0
	F0 Var	8636.1	4804.7	8047.0	4422.5	6319.8	6115.8	7333.6	4172.0	9012.8	3988.3	6376.6	5804.9
	Q25perc	7406.3	8252.5	7332.2	7531.6	7485.0	8544.7	7294.0	7530.3	8161.5	7445.0	7803.0	6977.1
	Q50perc	8221.8	9237.8	7953.1	8430.4	8094.5	9502.0	8100.4	8677.6	10155.3	8569.0	9449.9	9113.9
	Q75perc	9762.2	11469.7	10549.3	10405.1	9573.3	11227.9	11086.8	11027.2	14319.6	11181.6	13796.3	12628.6
	Fpeak	161.9	152.2	146.0	147.4	156.9	158.4	153.0	154.2	175.7	171.2	164.8	179.1
	percTime of max intensity	50.6	46.9	47.0	43.2	47.5	47.8	48.0	40.3	40.3	43.1	45.2	44.2
	jitter	0.030	0.026	0.035	0.026	0.020	0.023	0.030	0.027	0.019	0.024	0.023	0.034

4. Chapter

-

Trade-offs between vocal accommodation and individual recognisability in common marmoset vocalizations

Y. Zürcher[1,2], E.P. Willems[1], J.M. Burkart[1]

[1]Department of Anthropology, University of Zürich,
Winterthurerstrasse 190, Zürich

yvonne.zuercher@uzh.ch
0044 635 54 87

[2]Orcid-ID 0000-0001-6344-5339

Under review at Scientific Reports

4.1 Abstract

Recent studies find increasing evidence for vocal accommodation in nonhuman primates, indicating that this form of vocal learning is more prevalent than previously thought. However, high levels of convergent vocal accommodation (i.e. becoming more similar to partners) may compromise individual recognisability. We investigated in common marmoset monkeys (*Callithrix jacchus, n=20*) whether there are trade-offs between vocal accommodation to a new partner and the necessity to maintain individual recognisability. We quantified vocal accommodation in three call types over the process of pair formation and found that animals converged most in close contact calls (trill call), but less in calls where individual identity is more essential (phee- and food calls). In two out of three call types, the amount of accommodation was predicted by the initial vocal distance. Accommodation led to a group specific call signature in phee- and trill calls, but not in food calls. Moreover, it led to a drop in individual recognisability in in trill calls, but not in phee calls and food calls (where individual recognisability is more essential). Overall, our study shows that patterns of vocal accommodation vary between call types with different functions, suggesting trade-offs between vocal accommodation and individual recognisability in marmoset vocalizations.

Keywords: common marmosets, vocal accommodation, phee calls, trill calls, food calls, individuality

Acknowledgement: We would like to thank B. Kunz, A. Kissling, S. Falkner and R. Brügger for their help with processing the recordings. We are grateful to C. van Schaik and L. Vinicius for comments and suggestions to earlier versions of this manuscript.

This work was founded by the Swiss National Foundation (SNF) Grants 310030-13083 and 31003A-172979 (PI: J. Burkart).

Conflicts of interest: Each of the authors declare that they have no conflict of interest.

4.2 Introduction

Nonhuman primates hardly learn new call types, neither as infants nor as adults [Egnor and Hauser, 2004], and are thus often considered to lack vocal production learning altogether. However, according to the definition by Janik & Slater [Janik and Slater, 2000], vocal production learning occurs when vocal signals are modified due to the experience with those of another individual, and thus also includes the modification of calls already existing in the repertoire of an individual. This form of vocal learning, called vocal accommodation, is quite abundant in nonhuman primates, can be found in different species and contexts, and is functionally highly comparable with vocal accommodation in humans [Ruch et al., 2018]. Social vocal accommodation has been found in a variety of situations. Japanese macaques show increased vocal accommodation to higher ranking individuals [Sugiura, 1989], both Diana monkeys [Candiotti et al., 2012] and chimpanzees converge towards communication partners in a short time range [Mitani and Gros-Louis, 1998], and Campbell's monkeys share more similar call variations with individuals with whom they have stronger social bonds [Lemasson et al., 2011b]. Ample evidence for vocal learning in the form of vocal accommodation has also been reported in the callitrichidae, a primate family known for their vocal flexibility [Snowdon, 2001; Snowdon, 2013; Snowdon, 2017; Zhao et al., 2019]. In a colony of pygmy marmosets, the introduction of unfamiliar individuals lead to a shift in the call range of all the individuals [Elowson and Snowdon, 1994], and Wied's black-tufted-ear marmosets were found to modify their call structure after unfamiliar individuals were introduced into the colony room [Rukstalis et al., 2003]. In a recent study with common marmosets, individuals became more similar to the new colony after having been translocated from a colony with a different vocal variant (sometimes referred to as dialect) [Zürcher and Burkart, 2017], and the change of vocalization was most likely due to social vocal learning [Zürcher et al., 2019].

Vocal learning in the form of vocal accommodation is thus common in nonhuman primates (and quite likely also in other animals [Ruch et al., 2018]) and seems to serve a social function [Tyack, 2007]. Vocal convergence has been suggested to play a role in group cohesion, could lead to a group signature, facilitate recognition of group members, and is often linked to pair bond quality [Ruch et al., 2018; Tyack, 2007]. However, a hitherto neglected aspect of vocal convergence is that it may also have a less desired consequence: Converging towards a communication partner may reduce the individual recognisability of a vocalization. This will lead to a trade-off between the need to accommodate as a social signal, and the need to encode identity into calls. Such a trade-off predicts a "sweet spot", a specific vocal distance between communication partners that animals should aim for, where both the specific needs for accommodation and individuality are in equilibrium. Depending on call type and associated function, this equilibrium can be at a relatively larger or smaller vocal distance between communication partners.

This leads to a first set of predictions with regard to patterns of accommodation, for instance when unfamiliar individuals first meet to form a pair: First, the amount of convergence should differ between call types with different functions. Convergence can be larger in call types for which individuality is less important, such as close distance calls, which are given when individuals are in visual contact and acoustic information about individual identity is thus redundant. In call types for which transmitting individuality is paramount, however, as for instance in long-distance contact calls that are given out of visual contact, convergence should be constrained by the necessity to maintain individual recognisability. Thus, overall

we expect more vocal convergence in calls that do not have to transmit individual identity. Furthermore, if there is an optimal vocal distance between partners, a "sweet spot" that optimally combines the social advantages of convergence while maintaining individual recognisability, the amount and direction of accommodation should depend on the initial vocal distance between partners. Thus, pairs whose initial vocal distance is larger than the optimal distance should converge over the time of pair formation, whereas pairs who happen to have vocalizations that are more similar than the optimum should diverge. Consistent with this idea, Snowdon and Elowson found vocal convergence in pygmy marmosets in only 3 out of 4 newly formed breeding pairs; the individuals of the fourth pair who did not converge were very similar to each other already before pair formation [Snowdon and Elowson, 1999].

A second set of predictions following from a hypothesised trade-off between accommodation and individuality concerns the impact of accommodation on group identity and individual recognisability. In particular, accommodation should differently impact the emergence of a pair signature and individuality depending on call function. In call types such as long distance contact calls for which transmitting identity is essential, animals should converge little and in particular not at the expense of call individuality. This may or may not lead to a pair signature in such call types. In short distance contact calls however, where individual identity is less important because callers are typically within visual contact, there should be more flexibility to engage in accommodation and develop a pair signature, which can be achieved at the expense of individual recognisability. Table 4.1 provides an overview over these predictions.

To investigate potential trade-offs between accommodation and individuality, we measured the vocal output of a total of 20 common marmosets in 10 newly formed breeding pairs over the process of pair formation. Common marmosets are known to show a high degree of vocal flexibility and a certain degree of vocal learning (Snowdon 2013), including babbling in infants [Elowson et al., 1998a; Elowson et al., 1998b; Snowdon and Elowson, 2001], the importance of social input for vocal ontogeny, including feedback by parents [Koda et al., 2013; Takahashi et al., 2016; Takahashi et al., 2015; Takahashi et al., 2017] or acoustic differences between colonies or populations [Crockford et al., 2004; de la Torre and Snowdon, 2009; Mitani et al., 1992; Zürcher and Burkart, 2017]. We analysed three different call types with different social functions (phee calls, trill calls and food calls) that are regularly produced in a within-group context.

Table 4.1 Overview over the predictions: The table provides a short description of the three call types and how important transmitting identity (ID) was for each of them, followed by our expectations for each call types and the different predictions.

	Phee calls	**Trill calls**	**Food calls**
Call description	Long distance contact calls, produced when separated, out of sight	Close distance contact calls, usually produced within sight of others	Communicates detection and intent to share food; with or without close contact
Importance of ID	Crucial, as animals have to convey their ID	Less important; redundant as ID is directly visible	Important for receiver to know which individual is willing to share food
Predictions: Patterns of accommodation			
Extent of accommodation	Little accommodation expected	Highest amount of accommodation expected	Little accommodation expected
Accommodation correlates with initial vocal distance	Expected	Expected	Expected
Predictions: Impact on group identity and individual recognisability			
Emergence of *group ID* after pair formation	Expected (based on literature)	Expected	No *a priori* expectation
Reduction in individual recognisability	Not expected	Possible	Not expected

Phee calls are predominantly used as long distance isolation calls when individuals are separated from their mates or social group [Bezerra and Souto, 2008; Liao et al., 2018]. Phee calls are known to encode individual identity as well as group identity and sex [Jones et al., 1993; Miller et al., 2010; Norcross et al., 1999]. They typically elicit answering phee calls from other group members or mates, and animals often engage in turn taking, i.e. calling back and forth over several turns [Takahashi et al., 2013]. Common marmosets also use phee calls as a vocal territory advertisement [Norcross and Newman, 1997], although the social situation seems not necessarily to be encoded in the call [Miller et al., 2010]. Overall, these functions suggest that individual recognisability is essential in phee calls.

Trill calls are close distance social calls that are often exchanged between social partners in a very relaxed social situation. Wild common marmosets often produce trill calls in situations such as foraging or resting [Bezerra and Souto, 2008]. A study by Liao et al. could show that captive marmosets produce trill calls more often when being closer to their social partner and with a lower heart rate, so at a more relaxed state, compared to when producing phee calls [Liao et al., 2018]. Since trill calls are typically given from a close distance, we assume that individual recognisability is less essential than in phee calls.

The third call type we looked at were food calls (sometimes also referred to as chirp calls) [Bezerra and Souto, 2008]. Food calls are usually produced upon the detection of high value food and often indicate the willingness of the caller to share the food with other group members [Vitale et al., 2003]. Food calls are usually produced in bouts, and are given from variable distances. They seem to be more variable than trill- and phee calls, and might have some elements that are referential with regard to food type [Rogers et al., 2018]. Even though the main function of food calls is to announce the presence of food rather than signal individuality, the latter might still be important in situations where food calls are given from lager distances or out of sight, and could help receivers to correctly identify the animal currently willing to share food.

We first investigated the pattern of accommodation across call types. We quantified the amount of accommodation (both convergence, i.e. becoming more similar, and divergence, i.e. becoming more different) for each pair and each call type. We expected that the amount of accommodation would differ between call types. In particular, we expected more convergence towards partners in trill calls compared to phee calls and, to a lesser extent, food calls, because individual recognition of the caller is essential in phee and perhaps food calls, but not in trill calls.

Next, we tested whether the amount of accommodation was correlated with vocal distance prior to pair formation for each call type. We expected that the amount of accommodation would correlate with the initial vocal distance prior to pair formation, in particular when both convergent and divergent accommodation is present. Such a correlation would suggest the presence of an optimal equilibrium, a "sweet spot" between signalling similarity vs. individual identity, which is achieved by flexibly increasing or decreasing vocal distance to a partner.

In a next set of analyses, we quantified the impact of accommodation for group identity and individual recognisability. First, we tested whether individual calls could be correctly assigned to the corresponding pair. We expected such an assignment would only be possible after pair formation. Before pair formation, the animals did not have any social experience with each other, and it would be highly unlikely that they are more similar to their future social partner than to random individuals. Due to the lack of similarity between partners, assigning calls to a pair should not be higher than expected by chance. After pair formation, a pair signature can emerge due to vocal accommodation, as has been reported previously for phee calls [Miller et al., 2010]. Finally, we tested how accommodation and group ID had impacted individual recognisability. We expected that it should not impact individuality in call types for which individuality is essential (phee- and food calls), but could do so in trill calls.

4.3 Methods

4.3.1 Subjects

We recorded the vocal behaviour of 20 captive common marmosets over the process of pair formation of newly formed breeding pairs. All animals lived with at least one family member or a former partner until shortly before the introduction to their new breeding partner. Animals ranged from between 2 years to 9 years, and all individuals were unfamiliar with their new partner before the start of the study.

The enclosure of each pair measured 2.4 m in height × 1.5 m in depth × 0.8 m in width and was structured with branches, ropes, tubes and other enrichment material. All animals were fed twice a day (vitamin enriched mush in the morning and a mix of fruits and vegetables around midday) and in addition received different kind of animal or insect protein and / or gum once to twice a day. Water was always available ad libitum. The animals had regular access to spacious outdoor enclosures as well as to an additional testing room.

4.3.2 Recording procedure

The animals were recorded both before and during pair formation in a variety of different situations to elicit a broad range of calls covering a large part of the naturally occurring call spectrum of the marmoset (presentation of food to elicit food calls, recordings with partner to elicit trill calls, recordings when away from the partner to elicit phee calls). Before pair formation, individuals were recorded on several days over two to three weeks in their home enclosure either with a family member present or after being separated from their family group, as closely in time to pair formation as possible. After pair formation, we recorded the animals in different conditions, where they were either alone or together with their partner, on one to three days a week up to 13 weeks after pair formation. We recorded them both in either their home enclosure or in an additional, familiar experimental room which was connected to the home enclosure by a system of tubes through which the animals could walk. When recorded in their home enclosure, both animals of the pair were present. When recorded in the additional testing room, animals were either both present or they were separated from each other (either with the other animal still in the room with acoustic contact, or with the other animal back in the home enclosure) for up to five minutes. Both in the home enclosure and the test room, animals were recorded with or without highly preferred food (a mixture of mealworms, cashew seeds and nut-cookies). Recording sessions lasted between 20 – 30 minutes. During the recording, the experimenter was present in the room and pointed the handheld microphone in the direction of the focal animal, which changed every five minutes. The identity of the caller was directly annotated to the recording by the experimenter in real time using the labelling function provided by the AviSoft Recorder software.

Even though we tried to elicit calls from the animals, data recording remained largely opportunistic. Therefore, we do not have all call types of all the individuals over the whole time period. Pairs with less than a minimum of five calls per call type and per point in time where therefore excluded from further analysis, which led to a final sample of 8 – 9 pairs, depending on the call type.

4.3.3 Recording processing

The recordings were visually inspected in AviSoft Pro and each call saved as a separate file. We inspected and measured each call with the software Praat and extracted 15 (phee, food call) or 17 (trill) parameters per call ([Boersma and Weenink, 2009]; after a skript by [Briefer and McElligot, 2011]). We restricted our measurement to the fundamental frequency and extracted the frequency both the beginning and the end of the call, further the mean, minimal and maximal F0, the percentage of the call duration for which F0 was at the max, the absolute slope of F0, the mean variation of F0 per second, the frequency values at the first, second and third quartiles of energy, the highest frequency of the whole spectrum, percentage of time this highest frequency is reached and jitter, as well as frequency modulation rate and frequency

modulation extent for trill calls (see [Zürcher and Burkart, 2017] for a detailed description of the parameters). Calls were excluded from the final sample if there was background noise, if they overlapped with any other call or we could not measure the whole call correctly in Praat.

4.3.4 Statistics

We first performed principal component analyses for the three call types, based on the z-transformed values of the measured call parameters. We extracted all components with an Eigenvalue greater than the 95% quantile value obtained from 10000 datasets that were randomly generated and equal in sample size and dimensionality to our empirical data (Parallel analysis). For all further analyses, we used the PC-Factors extracted by this method. For factor loadings please see Supplementary Material Table S4.3.

4.3.4.1 Patterns of accommodation

To quantify convergence and divergence, we calculated the vocal distance between partners before the start of pair formation (bpf) and after pair formation (apf) for each call type (see Table 4.2 for the specific time after pair formation the apf – calls were recorded per pair and call type). We calculated the Euclidian distance between each call of the male and each call of the female within a pair based on the extracted PC-factors. It is important to note here that – as each call served as a reference for multiple distance measurements (each call was compared to each call of the partner) – these distance measurements between partners are not independent, and this non-independence has to be taken into account in the analysis. To estimate whether the vocal distance increased or decreased over time in the different pairs, we compared the distance matrix bpf with the distance matrix apf with a bootstrapped (taking into account the dependencies in the data) Welch t-test and calculated non-parametric 95% confidence intervals around the effect size to assess whether there was a significant change in the vocal distance. An increase in distance would indicate vocal divergence, a decrease in distance vocal convergence. We used the average of the Euclidian distances as a proxy for average vocal distance between partners for either point in time. The amount of accommodation was calculated as the change in vocal distance bpf to apf by subtracting the average vocal distance apf from the average vocal distance bpf. We calculated Pearson's correlation coefficients to test if the initial distance between pair mates and the amount of accommodation was correlated, separately for each call type.

4.3.4.2 Impact on group ID and individual recognisability

To test to what extent calls could be correctly assigned to the pair bpf and apf, we performed separate permutated Discriminant Function Analysis (pDFA) [Mundry and Sommer, 2007] for the data from bpf and apf separately, using a script provided by R. Mundry, controlling for individual identity as well as sex. Likewise, to investigate whether accommodation had an impact on the individuality encoded in the calls, we performed a Discriminant Function Analysis (DFA) both before and after pair formation to quantify to what extent calls could be correctly assigned to the individual producing them, using the total of the correctly assigned calls as a measure of individual recognisability of calls. For both pair and individual identification potential, we and tested whether these proportions of correctly assigned calls changed from bpf to apf with a chi^2-test.

All analyses were performed in R 3.5.3, except for the chi^2-test for which we used the tools provided on http://www.quantpsy.org/chisq/chisq.htm.

4.4 Results

4.4.1 Patterns of accommodation across call types

To disentangle how the calls changed over time, we quantified the amount of accommodation (both convergence and divergence) for each pair and each call type. We found that for phee calls, 5 out of 8 pairs showed a significant amount of accommodation, of which 1 pair diverged and 4 pairs converged. In trill calls, 6 out of 9 pairs showed a significant amount of accommodation, all of which converged. In food calls, all 9 pairs showed a significant amount of accommodation, and 3 pairs converged, while 6 pairs diverged (see Table 4.2 and Figure 4.1). Convergence was thus most prevalent in trill calls (66.67% of all pairs), followed by phee calls (50%) and food calls (33.33%).

Table 4.2: Amount of accommodation (convergence and divergence) for each pair and call type. *Week* refers to the week after pair formation when the recordings for the "after"-comparison were made (for phee - / trill - / and food calls respectively). *α – level* gives the level at which the vocal distance was significantly different before and after pair formation (ns indicates that the change in distance was not significant). *t* indicates the effect size, and *Δ distance* the difference between the average call distance before and after pair formation. ***Positive Δ values*** indicate convergence, i.e. that the pair became more similar, ***negative Δ values*** indicate divergence.

Pair	Week last rec. (phee / trill / food)	Phee call			Trill call			Food call		
		α - level	t	Δ distance	α - level	t	Δ distance	α - level	t	Δ distance
WiscoNaut	10 – 11 / 9 / 11	.05	0.131	0.155	.001	0.229	0.367	.01	0.314	0.554
WashGatto	10 – 12 / 9 / 10	ns	0.136	0.13	.001	0.141	0.238	.01	0.294	-0.416
NikPuk	10 / 9 – 10 / -	.05	0.148	0.229	.001	0.254	0.293	-	-	-
MibbCon	9 – 10 / 7 – 9 / 9	.05	0.322	-0.528	.01	0.004	0.007	.05	0.071	0.083
LilCrak	9 – 13 / 9 / 9	.05	0.058	0.08	ns	0.039	-0.071	.01	0.165	-0.198
NalaTam	6 - 7 / 6 / 6	ns	0.019	0.042	ns	0.010	-0.013	.01	0.353	0.405
LeaKyr	6 – 9 / 7 / 9	.05	0.148	0.26	.001	0.318	0.391	.01	0.204	-0.434
MiaNari	10 / 7 / 10	-	-	-	.05	0.101	0.244	.01	0.110	-0.202
TogaMio	10 / 7 / 9	ns	0.064	0.119	ns	0.035	0.041	.01	0.194	-0.257
JajaMembo	- / - / 9	-	-	-	-	-	-	.01	0.184	-0.369

Next, we tested whether the amount of accommodation was correlated with the initial vocal distance of the individuals before pair formation, which would suggest that the individuals flexibly increased or decreased their similarity to reach an optimal equilibrium between signalling similarity vs. individual recognisability. While in phee calls the amount of accommodation was independent of the initial call distance (N = 8, Pearson's correlation coefficient = 0.381, p = 0.352), both trill calls (N = 9, Pearson's correlation coefficient = 0.744, p = 0.022) and food calls (N = 9, Pearson's correlation coefficient = 0.782,

p = 0.013) showed a positive correlation between the initial vocal distance and the amount of vocal accommodation (see Figure 4.1).

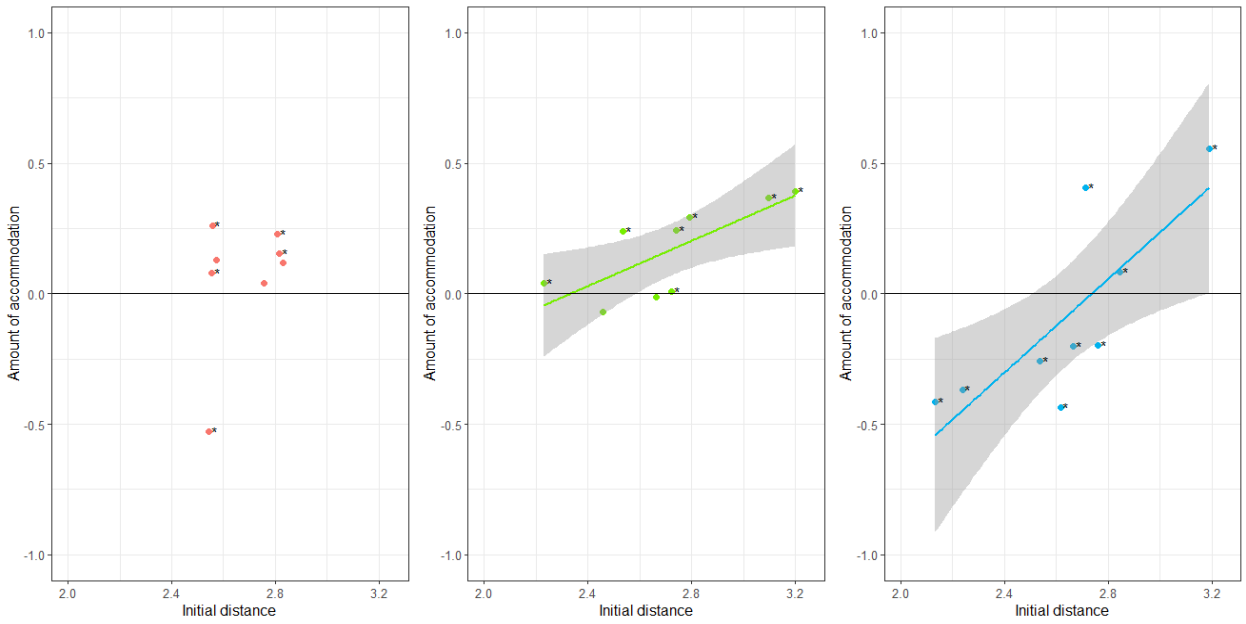

Figure 4.1: Relationship between the amount of accommodation and the initial vocal distance of each pair for phee-, trill- and food calls. ***Positive accommodation*** indicates convergence, i.e. that the pair became more similar, ***negative accommodation*** indicates divergence. In trill and food calls, the amount of accommodation is predicted by the initial distance.

4.4.2 Impact of accommodation on group identity and individual recognisability

To quantify the impact of the observed pattern of accommodation on group identity and individual recognisability, we compared changes in correct classifications of calls to group and individual before and after pair formation.

With regard to group identity, the calls from bpf could not be correctly classified to the future pair above random chance, although there was a trend for the food calls (pDFA; phee before: expected: 32.9%, correct: 35.3%, p =0.3, N = 1221; Trill before: expected: 31.1%, correct: 33.4%, p = 0.321, N = 474; Food calls before: expected: 21.1%, correct: 25.8%, p = 0.058, N = 3708). After pair formation, correct assignment of calls to the corresponding pair was higher than expected by chance for both phee- and trill calls (phee after: expected: 34.8%, correct: 43.51%, p = 0.039, N = 1407; trill after: expected: 35.4% correct 44.9%, p = 0.011, N = 1250). In food calls, there was still a trend for correct assignment, with the amount of correctly assigned calls similar to the value from bpf and lower than the values for phee- and trill calls apf (Food calls after: expected: 21.44%, correct: 25.7%, p =0.083, N = 3234). Correct classification thus increased for both phee calls and trill calls (chi^2 – test: Phee: $\chi^2 = 22.1$, df = 1, $p < 0.001$; Trill: $\chi^2 = 20.4$, df = 1, $p < 0.001$), but remained stable for food calls ($\chi^2 = 0.9$, df = 1, p = 0.34) (see Figure 4.2).

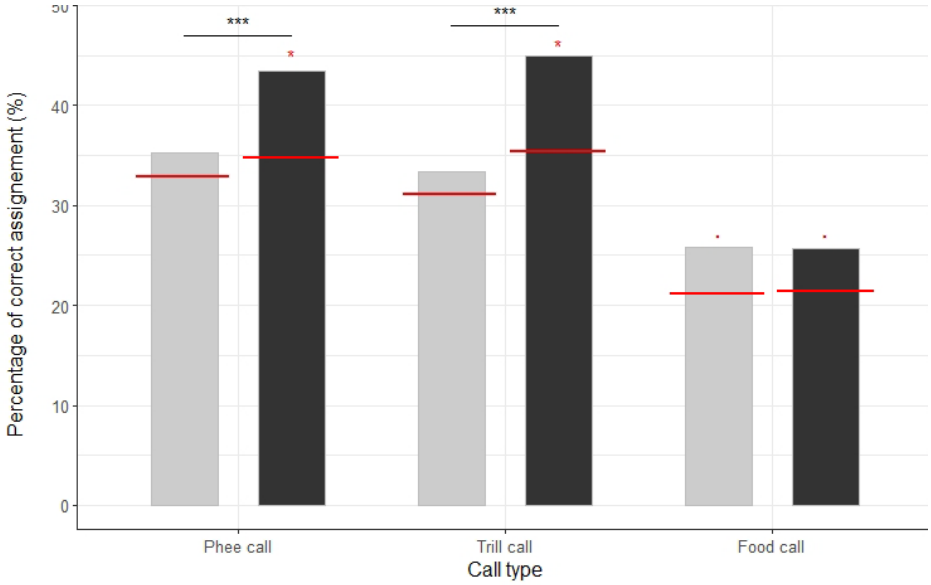

Figure 4.2: Classification of calls to pair. Before pair formation (light grey bar), the calls could not be assigned to the prospective pair better than expected by chance (red lines). After pair formation (dark grey bar), correct classification was significantly higher than expected by chance for phee calls and trill calls (red asterisks indicate significant better classification than expected), which indicates that a pair signature emerged in these two call types. The correct assignment of food calls did not change before and after pair formation and showed a trend before and after pair formation. The percentages of correct assignment were obtained by performing a permutated Discriminant Function Analysis. Correct assignment was significantly higher after pair formation compared to before, in phee calls and trill calls but not in food calls (chi^2 test, indicated by black asterisk)

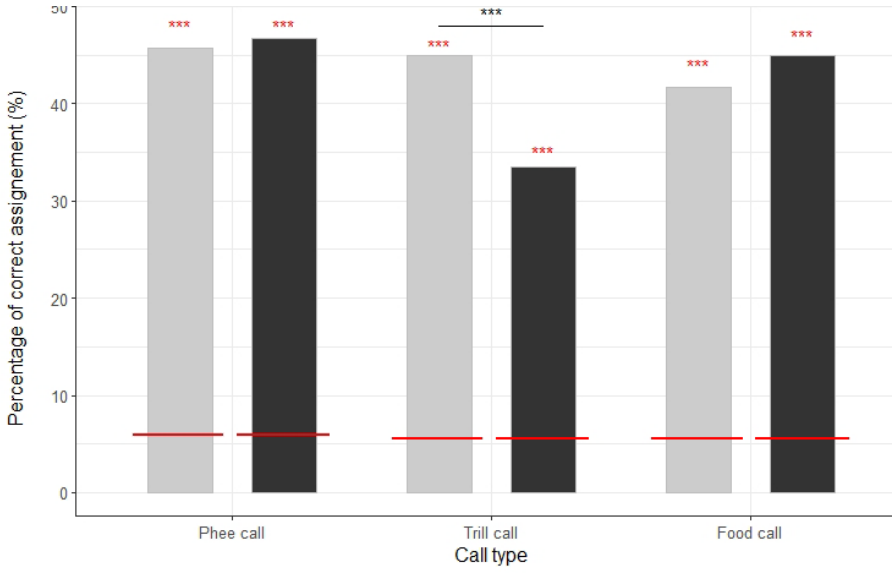

Figure 4.3: Individual recognisability. Percentage of correct assignments obtained from a discriminant function analysis. Calls can be attributed to the correct individual by discriminant function analysis significantly better than expected by chance (red line, indicated by red asterisk) in all conditions. The amount of correct assignment though significantly decreased in trill calls after pair formation (chi^2-Test, indicated by black asterisk). We did not observe a change in the level of correct assignment in phee calls and food calls. Light grey bars indicate values of correct assignment before pair formation, dark grey bars after

Individual recognisability, finally, was always above chance. The expected amount of correct classification by chance was around 6 % for each call type. When performing a discriminant function analysis, individual recognisability remained at comparable levels before (45.7%) and after (46.7%) pair formation for phee calls (chi^2 – test: χ^2=0.5, df = 1, p = 0.49) and food calls (41.7% bpf to 45% correct assignment apf; chi^2 – test: χ^2=3, df = 1, p = 0.1). In trill calls, however, individual recognisability significantly dropped from 45% (before) to 33.5% after pair formation (chi^2 – test: χ^2=19.5, df = 1, p > 0.001, see Figure 4.3).

4.5 Discussion

Increasing evidence for vocal accommodation in nonhuman primates has received a lot of attention in recent research because it suggests more vocal learning than previously assumed. When vocally accommodating, animals modify their vocalizations due to a social template, following the definition of vocal learning by Janik and Slater [1997]. Vocal accommodation often seems to serve a social function, reflecting social distance or the strength of a social bond. Nevertheless, an excess in vocal convergence can have disadvantages, when increasing vocal similarity leads to a loss in individual recognisability [Thomsen et al., 2019]. In this study, we explored potential trade-offs between the benefits of convergence vs maintaining individuality in call structure in common marmosets. To do so, we tested newly formed pairs and compared their vocalizations before and after pair formation. This situation has elicited vocal accommodation in pygmy marmosets previously [Snowdon and Elowson, 1999], but so far it was unclear whether and how marmosets would deal with the different requirements of converging to a partner while keeping their identity encoded in the calls. In this study, we therefore investigated how common marmosets accommodate to their partners in three different call types that critically differ in their function: Phee calls, which are long distance contact calls mainly produced when animals are separated from social partners; trill calls, which are close distance calls usually produced in close proximity, and food calls, which are emitted when animals find preferred food, often indicating willingness to share. In a second step, we examined to what extent their pattern of accommodation impacted pair- and individual call signatures, and how this was related to the different call functions.

4.5.1 Patterns of accommodation across call types

In our first set of predictions, we expected that the amount of convergence should differ between call types with different functions if there is a trade-off between the social function of accommodation and individual identity. We found vocal accommodation in all three call types, but to a different degree. Most animals converged on the trill call, and less in phee- and food calls, following our predictions. Whereas in phee calls and food calls we found both convergence and divergence, in trill calls we only found convergence. We further hypothesized that a trade-off between accommodation and individual identity would lead to a call-type specific “sweet-spot”, where both patterns are in equilibrium. This would lead to a different amount of accommodation depending on the initial vocal distance between future pair members, and would lead to convergence when the distance is larger than the optimum, and to divergence when the distance is smaller than the optimum. We found that in trill calls and food calls, the amount of accommodation was correlated with the initial vocal distance between pairs, but not in phee calls. From our data, we cannot conclude if this correlation is absent in the phee calls, or if this is an artefact of the rather small sample size. For trill calls, pairs with an initially larger vocal distance converged

more than pairs with an already smaller distance and in food calls, the correlation between the initial distance and the amount of accommodation indicates that pairs with an initially larger vocal distance converged, while pairs with an initially small vocal distance diverged. Both results are consistent with the hypothesis of a "sweet spot", an optimal vocal distance between partners and the hypothesis of a trade-off between social accommodation and preserving individual identity in calls, leading to a different pattern of accommodation depending on the call function and initial distance. Our next step was to investigate whether these differences in accommodation pattern also led to a difference in the change of group as well as individual ID encoded in the different call types.

4.5.2 Impact of accommodation for group identity and individual recognisability

In the second set of predictions, we assumed that vocal accommodation should impact the formation of a pair signature as well as potential changes in individual recognisability differently depending on the call function. In long distance contact calls (phee calls) where transmitting the identity of the caller is essential, we expected to find no decrease in the individual recognisability of vocalizations, whereas in close contact distance calls a such a decrease would be possible without detrimental consequences.

Both phee calls and trill calls could be attributed to the correct pair significantly more often after pair formation than before, indicating that the calls of the two pair mates became more similar in structure during pair formation. This is arguably the result of the observed convergence in some or most of these pairs, whereas the lacking group signature in food calls would result from the high proportion of divergence in these calls. Group signatures are quite common in different animal species including bats [Boughman, 1997; Boughman and Wilkinson, 1998; Pearl and Fenton, 1996], cetaceans [Filatova et al., 2007; Ford, 1991; Rendell and Whitehead, 2003; Smolker and Pepper, 1999; Strager, 1995] as well as in many birds [Henry et al., 2015]. They can serve different purposes, but often indicate that a certain individual belongs to a group. This facilitates the recognition of group members [Tyack, 2007], but can also play a role in inter-group encounters. Especially when group encounters serve to explore potential mating opportunities, it might be beneficial to infer group identity from calls. Since phee calls are often involved in initiating inter-group encounters, and appear to contain information about social status (breeding vs. non-breeding animal) [Norcross and Newman, 1997; Norcross et al., 1999], they appear particularly suitable to gather information about potential mates in other groups.

In trill calls, which showed the highest level of convergence, we found a significant decrease in the individuality of the calls (calls could be assigned to the correct individual less reliably). In phee calls and food calls, the individuality did not change even though convergence occurred in some pairs. Again, these two results are consistent with the hypothesized trade-off between vocal accommodation and individual distance, and how this trade-off can affect calls differently depending on their function.

Whereas convergence occurred in all three call types, divergence occurred in food calls in particular. The food calls of the future pairs appeared to be very similar already before pair formation (they could be assigned to the future pair almost better than expected by chance), which arguably led to this high level of divergence. It thus appears that individual recognisability is indeed important for food calls, and future studies using playbacks will help disentangle why this is the case. What we did not consider in this study is the fact that food calls are normally produced in call bouts that contain several individual food call

elements. In our analysis, we only analysed the single elements but not the information that is potentially encoded in the call bout. An intriguing possibility is that marmosets also accommodate to their partner with regard to bout structure (e.g. duration, number of elements), similar to the occurrence of accommodation in humans at multiple levels, from acoustic structure to word choice and syntax [Ruch et al., 2018]. Moreover, some elements of marmoset food calls appear to be functionally referential [Rogers et al., 2018]. Taken together, the food calls thus appear more heterogeneous than the other two call types analysed here, and additional studies will be necessary to fully understand how they change together with changes in social context.

Whether or how vocal similarity or dissimilarity is used as a social signal in common marmosets is still an open question. Our main research focus of this study was to investigate how the different needs for accommodation and individuality can be accounted for. Based on studies in other animals, it is well possible that accommodation, or another mean of vocal flexibility, is used by common marmosets to signal or maintain their pair bond [Ruch et al., 2018]. Our results suggest that trill calls are particularly likely candidate vocalisations for such a function, as they are more prone to accommodation and appear less constrained by the need to maintain individual recognisability. Moreover, they are often produced by animals which are in close contact and have a strong social bond [Liao et al., 2018]. To investigate how common marmosets perceive similarity in calls, playback studies will be the most informative next step. Presenting playbacks that simulate pairs with more or less similar calls, could answer the question if and how common marmosets use potential information encoded in different call types.

Vocal learning is still considered rare in nonhuman primates [Egnor and Hauser, 2004]. In this study, we could confirm that common marmosets engage in vocal accommodation – a form of vocal learning – quite regularly – but also, that they face trade-offs between similarity and individuality. Together, this corroborates that common marmosets have a high level of vocal flexibility, and that they use vocal accommodation as a very flexible system which might get regulated differently depending on call types and call type functions.

4.6 Supplementary Material

Table S4.3: This table provides an overview over the factor loading of the different parameters on each PC-Factor for each call type separately.

	Phee call				
	PC1	PC2	PC3	PC4	PC5
sound_duration	0.10	-0.47	0.18	0.09	-0.19
MeanF0	-0.41	-0.11	-0.05	0.13	-0.05
F0start	-0.31	-0.24	-0.37	-0.20	0.02
F0end	-0.35	-0.08	0.19	0.30	0.11
MaxF0	-0.39	-0.06	0.08	0.28	-0.07
TimeMaxF0	-0.02	-0.16	0.60	0.34	0.33
MinF0	-0.33	-0.26	-0.29	-0.13	0.10
F0absslope	-0.19	0.52	-0.05	0.17	-0.11
F0Var	-0.19	0.51	-0.06	0.19	-0.11
Q25	-0.38	-0.10	0.06	-0.12	-0.11
Q50	-0.34	0.10	0.25	-0.27	-0.01
Q75	-0.09	0.20	0.39	-0.52	0.12
Fpeak	0.03	-0.11	0.09	-0.10	-0.48
Timeofmaxintensity	-0.09	0.01	0.27	-0.44	-0.10
jitter	-0.02	0.03	-0.17	-0.12	0.73
Standard deviation	2.32	1.61	1.18	1.15	1.10
Proportion of Variance explained	0.36	0.17	0.09	0.09	0.08
Eigenvalue	5.41	2.61	1.39	1.33	1.20
	Food call				
	PC1	PC2	PC3	PC4	PC5
sound_duration	-0.12	-0.19	0.07	-0.28	0.56
MeanF0	-0.41	-0.09	-0.07	0.08	-0.01
F0start	-0.41	0.07	-0.12	0.09	0.04
F0end	-0.36	-0.33	-0.04	0.04	-0.10
MaxF0	-0.41	0.06	-0.11	0.10	0.05
TimeMaxF0	0.08	-0.23	0.41	0.47	0.13
MinF0	-0.36	-0.33	-0.04	0.03	-0.10
F0absslope	-0.15	0.50	-0.07	0.33	0.17
F0Var	-0.13	0.50	-0.30	-0.08	-0.25
Q25	-0.36	0.06	0.21	-0.05	0.05
Q50	-0.21	0.27	0.53	-0.10	-0.02
Q75	-0.07	0.25	0.58	-0.28	-0.12
Fpeak	-0.04	0.03	-0.10	-0.20	-0.21
Timeofmaxintensity	0.03	-0.09	0.17	0.47	-0.54

jitter	0.05	0.20	-0.04	0.46	0.45
Standard deviation	2.37	1.42	1.25	1.22	1.08
Proportion of Variance explained	0.37	0.13	0.10	0.10	0.08
Eigenvalue	5.62	2.01	1.57	1.50	1.16
			Trill calls		
	PC1	PC2	PC3	PC4	PC5
sound_duration	-0.18	-0.05	-0.30	-0.48	0.09
MeanF0	-0.43	0.11	0.05	-0.02	-0.03
F0start	-0.37	0.22	0.20	0.10	-0.13
F0end	-0.43	-0.01	0.01	0.00	0.13
MaxF0	-0.44	-0.04	0.03	0.03	0.04
TimeMaxF0	-0.16	-0.06	-0.23	-0.29	0.56
MinF0	-0.38	0.24	0.16	0.05	-0.07
F0absslope	-0.07	-0.43	0.09	0.42	0.07
F0Var	-0.16	-0.45	-0.25	0.02	-0.13
Q25	-0.14	-0.18	0.50	-0.19	-0.17
Q50	0.13	-0.29	0.44	-0.33	-0.12
Q75	0.03	-0.22	0.36	-0.36	0.19
Fpeak	-0.02	-0.01	-0.11	-0.28	-0.19
Timeofmaxintensity	-0.07	-0.17	-0.04	-0.01	0.21
jitter	0.11	0.07	0.20	0.06	0.57
F0.Var	-0.12	-0.45	0.01	0.33	0.13
Fm.Rate	0.08	0.28	0.30	0.18	0.34
Standard deviation	2.26	1.69	1.29	1.26	1.06
Proportion of Variance explained	0.30	0.17	0.10	0.09	0.07
Eigenvalue	5.12	2.85	1.67	1.58	1.12

5. Chapter

-

General Discussion and Conclusions

5.1 Summary of findings

During my PhD, I found that common marmosets have vocal dialects that are most likely the cause of vocal production learning in the form of vocal accommodation (chapter 2 & 3), and that accommodation is variable and depending on call type and call function (chapter 4). In my first study, I could show that common marmosets exhibit acoustic differences between three different, isolated captive populations. As vocal differences can occur due to different reasons, I wanted to investigate if vocalizations were flexible, and whether vocal dialects were most likely explained by genetic differences, environmental differences between the populations, or by patterns of social learning. I therefore translocated animals between two colonies as well as between different physical environments and tracked their vocalizations through the whole translocation process. I found that a change in the physical environment did not lead to long-lasting changes in vocalizations, but integration into a new colony did. In both phee and trill calls, the translocated marmoset became more similar to the dialect of the new colony over the duration of 16 weeks. These results indicate that vocal differences between populations are best explained by vocal production learning rather than by genetic or environmental differences. To understand the process of accommodation in more details, I studied vocal accommodation on a dyadic level. I formed ten new breeding pairs and recorded the vocal development of each individual over a time of six to thirteen weeks. I found that common marmosets regularly accommodate to their partner, but the amount and degree of accommodation was depending on the initial distance and the call function. In calls that served mainly for individual identification (phee calls) or where individual identification was important (food calls), I found a less clear pattern of convergence, respectively a higher amount of divergence. Only in trill calls, where identity can most probably be determined by other means than call characteristics, I exclusively found convergence between the pair partners. These results showed that vocal accommodation is not simply a rigid reaction taking place under any circumstances, but seems to be rather flexible and might be regulated differently in different call types with different functions.

Both in accommodation to a new social group as well as during pair formation, I found a similar pattern of accommodation depending on call types. Trill calls were in both studies the call type that showed the strongest accommodation, followed by phee calls. In both studies, food calls showed more divergence than the other two call types, indicating that animals were aiming for a larger vocal distance or more individuality in those calls. As food calls can also be produced when animals are very close to each other or even simultaneously with other animals, it might be more difficult for a receiver to distinguish between potential callers, which makes it even more important that calls are easily distinguishable. The divergence I observed in this call type might therefore have helped to keep the individual distinctness of the food calls. Further, the fact that food calls might be referential to a certain degree and are usually produced in bouts makes them more diverse than the other two call types analysed, which may partially explain the difference I found between the call types.

5.2 The function of vocal accommodation

In Ruch et al. [2018], we argued that social vocal accommodation should occur "*(i)*...in species that also engage in environmental accommodation based on auditory–vocal feedback, and (*ii*) that vocal divergence should only be found in species that also show evidence for vocal convergence. [Ruch et al., 2018 pp. 13]". Marmosets fulfil the second prediction by showing both convergence and divergence, but I could not confirm the first prediction, as I found social, but not environmental accommodation in my setting. It might be interesting to investigate environmental accommodation focusing on environments that show a stronger differentiation in physical structure and therefore differ more strongly in attenuation

and reverberation, to see whether marmosets do adjust their calls to these features, as it has been found in female olive baboons [Ey et al., 2009]. Nevertheless, findings of a Lombard effect in common marmosets suggest that they do react on strong environmental noise, also over a longer period of time, therefore showing a form of environmental accommodation [Brumm, 2004; Zhao et al., 2019].

It is still an open question why common marmosets show this strong tendency to social vocal accommodation. Human language has – among other functions – been suggested to serve as replacement for time consuming grooming of social partner, and therefore to serve an affiliative function. The advantage of using vocalizations instead of grooming for social bonding consists of the possibility to communicate both over a larger distance, so animals do not have to be physically very close together, being able to bond with several individuals at once, as well as perform another task at the same time [Dunbar, 2003]. It is therefore very likely that vocalizations as well as vocal accommodation serve a function in social bonding in primates. Indeed, in the literature, vocal accommodation can often be linked to social bond strength or social partner preference. This function seems to be very old and highly comparable between humans and primates, and potentially many other gregarious species [Ruch et al., 2018]. Vocal similarity within groups or pairs might serve a variety of purposes. Most often, similarity within groups gets interpreted as a kind of group signal, like a group "batch" and to facilitate recognition of group members. Further, similarity might help creating a relaxed environment or context within which animals interact with each other. If animals show little similarity otherwise, creating similarity in the communicative domain, be it vocal or gestural, could provide access to the benefits that come from similarity with a partner [see also Roberts and Roberts, 2017].

Generally, humans as well as other animals seem to have a preference for others that are more similar to oneself (homophily). Human infants prefer agents with a similar vocal variation as themselves for cooperation partners [Haun and Over, 2015], and chimpanzees seem to become friends with individuals of similar personalities [Massen and Koski, 2014]. Similarly, also rhesus macaques seem to prefer partners with similar personalities, at least the females [Capitanio et al., 2017; Weinstein and Capitanio, 2008]. Similarity with one's partner seems to facilitate coordination and cooperation with each other, potentially due to increased trust into the partners behaviour [Massen and Koski, 2014] or due to an increased predictability of the partners behaviour and reactions [Berger and Calabrese, 1974; Parkinson et al., 2018]. This might be especially important for species that are highly reliant on cooperation partners, such as during cooperative hunting or cooperative breeding. Indeed, marmoset monkeys become more similar in personality not only with specific dyadic partners but on a group level, leading to a group personality [Koski and Burkart, 2015], supporting the importance in this species to build a foundation allowing successful cooperation with all group members.

5.3 Remaining open questions

The results of my studies nicely fall in line with what is known about common marmoset communication. The results about vocal accommodation highlight the high vocal flexibility this species shows in a broad range of contexts, and corroborate the important role of common marmosets in studying human language evolution. Still, there are some open questions that will be highly interesting to answer in further studies about vocal accommodation and vocal learning in common marmosets.

We do not know yet whether social behaviour is correlated to social accommodation on a long- or short-term base. This could be studied by testing for correlations between different behavioural patterns like affiliative or aggressive behaviours and acoustic parameters on different time scales, which could give us

evidence on what function vocal accommodation serves in common marmosets. Knowing whether and at what time scale vocal accommodation influences the behaviour of common marmosets could also be important when studying if vocal similarity has any effect on the fitness of a pair. We know from previous studies that social bonds are essential for common marmosets, as the strength of the pair bond has an influence on the quality of the infant care of the pair [Finkenwirth and Burkart, 2018]. In this situation, social accommodation could facilitate the establishment of a strong pair bond by either increasing the similarity between partners, which might make them more preferable towards each other or by providing a function of grooming at a distance, leading to a stronger social bond. As cooperative breeders, both mates are interested in a stable relationship. The males need some guarantee that they will sire the offspring of the breeding female, while the females need the guarantee that the male will still be around by the time the infants are born. Both male and female of a pair need to have a means to estimate bond quality [Snowdon, 2017]. Vocally accommodating to one's partner could therefore, together with other behaviours signalling pair bond, work as a signal to the partner that one is committed to the pair bond and the stability thereof. To test this, it would be essential to both test whether vocal similarity or accommodation has an influence on the pair bond and ultimately also on the individual fitness. In the studies that were part of my thesis, I did not test any link between the bond strength of the breeding pairs and the amount of vocal accommodation. I collected the behavioural data during pair formation though, and run some preliminary analyses. So far, I did not find any evidence for pair bond strength being linked to the strength or direction of accommodation, but this data need to be investigated in more detail still. In future analysis, I could potentially link the data on vocal similarity and vocal correlation with data on fitness, for example by measuring the time it takes a newly formed breeding pair to have their first set of infants, and whether this shows any correlation to the initial vocal distance, the amount of accommodation or the vocal correlation between partners.

The way I analysed the data, I cannot say which parameters were modified most by vocal accommodation, because I rather looked at the call structure as a whole. The change I found in Euclidian distance indicate that common marmosets modified several parameters in their calls simultaneously. Even though I found a combined effect, it could be interesting to analyse the changes in more details to get an understanding of how they were initially different and what exactly changed. Still, even if we can determine which parameters are changing most in the calls analysed here, we do not know if these changes are salient and important for marmosets. Experiments testing the reaction of marmosets on changes in calls would be necessary to investigate this.

In general, little is known about how marmosets perceive calls from other conspecifics, what information they gain by them and what their understanding of different calls is. There are some studies that give insight into the communicative understanding of marmosets though. Rukstalis and French [2005] found that playback of phee calls of a partner can reduce stress levels in a separated individual, which indicates that marmosets can identify the sender of the call. Also, a study along the same line could show that the playback of food calls could elicit a more quiet behaviour in a colony, indicating that the mood of the caller was contagious to a certain degree to the other animals in the colony [Watson et al., 2014]. Finally, a study by Rogers et al. [2018] suggests that food calls in marmosets are referential. We repeated this study and also tested whether marmosets were reacting accordingly to playback of food calls produced for different food types. Preliminary analysis suggest that marmosets do indeed understand the different signals of conspecifics, but the data needs to be further analysed before any conclusion can be drawn. Still, there is a lot to learn about the perceptive and cognitive abilities involved in marmoset communication. This will

be highly interesting considering that understanding calls made by others is potentially highly similar to the understanding of syntax [Seyfarth et al., 2005]. Until we invest more in the understanding of communication in primates, the gap in our knowledge about this is also a gap in our understanding of how this skill could have evolved in the human lineage.

5.4 About human language evolution – why no one else talks

5.4.1 Elements of language

As we could see already in the introduction, many elements of human language are also present in other animals, and most probably were present in the last common ancestor of human and other great apes. So what did it take that in one species a complex and highly flexible communication system evolved out of these building blocks? Human language consists of several core characteristics, like vocal learning, structural rules, as well as semantics, or meaning. Besides this structural and perceptual abilities, both sender and receiver need the cognitive capacity to be able to process the communicative content as well as the communicative signal [Tallerman and Gibson, 2012]. Further, human language is fundamentally about transmitting information or knowledge [Tomasello, 2010], therefore potential content plays an important role in communication. To allow for information transfer between individuals, potential communication partners need to be in a situation that facilitates the transmission of content. This contains several aspects: the individual that has information needs to have a certain desire to transmit this information, it needs a communication system that is honest and allows to transmit variable information and it needs a situation where the receiver knows that information is transmitted and therefore is attentive to the sender in a cooperative way [Fitch, 2005; Fitch, 2007]. I will discuss the aspects of communicative content, cognitive capacities and communicative attitude in the following three sections, investigating how these elements are present in humans and which of them can be found in chimpanzees as the closest living relative to humans as well as in marmosets, which share with humans a similar social system.

5.4.1.1 Content

Content, i.e. the information that is transmitted by a communicative signal, can be a variety of different things. All communication signals transmit some information, be it about the toxicity of a colourful butterfly, the mate quality of a bird singing, the location of a good food source in the honey bee dance or about the presence of a potential danger in alarm calls in many species [Liebal et al., 2014]. In each of this cases, the content is rather restricted to a specific topic and meaning, quite opposite to the variety of contexts humans use a signal for. In humans, communication can include exchanges of simple information to teaching of abstract techniques, making plans, telling stories or gossiping about other people [Dunbar, 2003; Tomasello, 2010; Tomasello, 2019]. Highly sophisticated culture, complex skills for food acquisition as well as extended and important social relationships are an almost limitless source for communicative content.

Chimpanzees also have rather complex foraging techniques (ranging from ant-fishing and nut-cracking to hunting monkeys and cracking turtle and crabs) and form strong social bonds [Koops et al., 2019; Massen and Koski, 2014; Pika et al., 2019; Whiten et al., 1999]. They show differences in behaviour between groups that are not explained by differences in the environment or genetics, but are based on social learning, and deemed to be cultural [Whiten et al., 1999]. Each of these contexts could be a source for communicative content that could be worth sharing with other group members.

Marmosets do not have complex foraging techniques, but do have to coordinate group activities such as infant transfers between caregivers, vigilance and territory defence or the selection of sleeping trees. They

also share food mainly with immature group members and probably even show teaching behaviour towards their infants. They often use vocal communication in these situations. Infant transfers in common marmosets are often accompanied by vocalizations of both the adults and the infants (personal obs.,). Territory defence is also a highly vocal behaviour, and marmosets use vocal signals both to advertise their territory as well as to recruit other group members and during group encounters [Bezerra and Souto, 2008; Lazaro-Perea, 2001]. Marmosets and tamarins are also highly vocal during foraging. They produce specific food calls when they find food they are willing to share [Bezerra and Souto, 2008; Roush and Snowdon, 2001; Vitale et al., 2003]. In common marmosets, food calls seem to be at least partially referential [Rogers et al., 2018] and reflect food preference in cotton top tamarins [Elowson et al., 1991]. With their communicative behaviour, cotton top tamarins seem to provide information about both good as well as unpalatable food to their immature and inexperienced group members by producing alarm calls when encountering unpalatable food [Snowdon and Boe, 2003]. Still, all of these contents are relatively simple. Callitrichids do not use and or manufacture tools, nor do they hunt in coordinated groups or have been reported to have cultures as such.

5.4.1.2 Cognition

The use of tools or other complex foraging techniques as well as the development of culture are largely restricted by an animal's cognitive abilities, with cognitive capacity being largely depending on brain size [Deaner et al., 2007; Reader et al., 2011]. Humans have by far larger brains, in relation to their body size and absolutely, than any other primate, and show many extraordinary abilities rooted in their strong cognitive capacities [Isler and van Schaik, 2014; Reader et al., 2011]. Also chimpanzees are highly capable when it comes to understanding the physical world. They are very proficient in understanding space, categories as well as quantities, and when it comes to tool use, chimpanzees are able to mentally represent the problem in order to find appropriate solutions [Tomasello, 2019]. Chimpanzees are able to understand intentional behaviour in other agents, as well as to understand what others see and hear, especially in competitive situations. When tested in a large cognitive battery testing a range of cognitive abilities, the performance of chimpanzees is comparable to that of two year old human children in tasks concerning the physical domain [Herrmann et al., 2007]. Concerning communication, we know that great apes have the cognitive capacity to learn and use language systems when in intense contact with human care givers. The so-called language trained apes are evidence that the lack of language in chimpanzees is not due to a lack of cognitive capacities to process language. Chimpanzees can use communicative signals with humans correctly [Gibson, 2012; Russell et al., 2005], and Kanzi, a language trained bonobo, showed a level of understanding grammatical rules like syntax that were comparable to those of a four year old child [Savage-Rumbaugh et al., 1993].

Marmosets, on the other hand, are very small primates, which naturally also have small brains and are therefore limited in their potential cognitive capacity. Indeed, they are also found to show rather low cognitive skills when compared to other primates [Deaner et al., 2006], but within the range of what would be expected when considering their brain size [Burkart and van Schaik, 2010]. What common marmosets are exceptionally good at are tasks in the socio-cognitive domain: despite their unimpressive general cognitive performance, marmosets outperform their (larger brained) sister species in the socio-cognitive domain (although potentially mainly due to an increased motivational interest and not due to a better cognitive ability per se [Burkart and van Schaik, 2010]). They are surprisingly good in social learning: common marmosets show true imitation in a copying task [Voelkl and Huber, 2000], only copy behaviour that they perceive as intentional [Burkart et al., 2012] and pay attention to the actions of other group members, especially when an actor is engaged in a problem solving task compared to an actor that is

merely exploring [Range and Huber, 2007]. They follow the gaze of an experimenter, also around a barrier [Burkart and Heschl, 2006], and discriminate between what others can see or not [Burkart and Heschl, 2007].
Interestingly, chimpanzees are not very proficient in the socio-cognitive domain, which is especially surprising as young children are very good at those already, and even outperform chimpanzees of the same age [Herrmann et al., 2007; Wobber et al., 2014]. Moreover, in contrast to human children, socio-cognitive skills of chimpanzees are at their strongest in a competitive context rather than in cooperative situations [Tomasello, 2019]. For example, chimpanzees seem to have difficulties following gaze direction cues to select a baited cup when the cue is cooperative, i.e. the experimenter is looking at the baited cup. Arguably, chimpanzees understand the looking behaviour of the experimenter more as a signal of food possession than as a signal of help, and therefore choose cups mainly at random levels. In line with this interpretation, chimpanzees were well able to use the cue in a situation where the experimenter indicates the empty cup instead, allowing the chimpanzee to select the baited cup [Tomasello, 2019].
Tomasello and colleagues concluded from this results that humans, contrary to chimpanzees, are not only skilled in competitive, but also in complex forms of cooperative behaviours [Tomasello et al., 2005]. A similar argument could be made for marmosets, which also show increased socio-cognitive skills. This increased socio-cognitive performance both in humans and in common marmosets might be an important reason for the elevated communicative behaviour found in those two species. I will discuss this in more details in the next section.

5.4.1.3 Communication attitude and cooperative mind set

Even if a species shows behaviours that lead to communicative content worth communicating about and possess the cognitive abilities to both use a complex communication system that allows transmitting the content as well as to understand the transmitted information, it still needs a situation that allows to transmit the information. In a first step, signals have to evolve that allow to transmit the content, and signals have to be honest, as information transmission only works with those [Fitch, 2000]. From an evolutionary point of view, communication is only beneficial if it increases the fitness (direct or indirect) of the individuals communicating. The evolution of real honest signals is relatively complicated though, as it can be easily hampered by free-riders. Honest signals are therefore hardly evolutionary stable strategies. As Fitch establishes in his review 2007, honest signals evolved most likely in a system with strong kin selection [Fitch, 2007]. In such a system, the evolution of honest signals is supported as it is in everyone's interest both not to deceive others and to pay attention to these signals instead of ignoring them. One requirement for the evolution of language is therefore a social, cooperative setup where honest signals can persist, which is most likely the case in a system with high kin selection [Fitch, 2007]. Individuals also need a certain proneness to prosocial behaviours and cooperation, an urge to share information and a prosocial attitude towards others that allows for close proximity and attention towards other individuals without tension. In a situation like this, individuals can share information freely and voluntarily [Burkart et al., 2018; van Schaik, 2016].

Cooperation in the form of food sharing is essential in traditional human societies. Food sharing within a group can prevent food shortages on days where a specific individual was less successful than others, which allows for a consistent food intake [Dyble et al., 2016]. This is crucial to allow a species to develop both short inter birth intervals while having slowly developing and costly infants. All in all, food sharing might be one of the fundamental behaviours that allowed for the increased human cognitive capacities by allowing the development and maintenance of big brains [Heldstab et al., 2019]. But humans do not

only share food, but also share information with each other. Human conversation is often an exchange of information, and humans share knowledge often without being prompted to, simply because they identify a piece of information to be of interest to their communication partner [Tomasello, 2010]. Sharing of information also includes teaching. Even if teaching in traditional hunter – gatherer societies is rarer than initially expected from a Eurocentric perspective, and occurs in different forms, it still occurs regularly and at higher frequency than in any other primate species, and also occurs already in children at relatively young age [Boesch et al., 2019; Boyette and Hewlett, 2018; Csibra and Gergely, 2011; Strauss and Ziv, 2012]. Especially cultural and social norms are transmitted via teaching [Salali et al., 2019]. Both the sharing of food and sharing of information is an expression of the prosocial and hyper-cooperative nature of humans, and is not found in other great apes at such an extent.

Humans develop elaborate forms of joint attention and perspective taking already very early during infancy. These abilities are both essential to establish joint goals and actions as well as in coordinating groups to achieve them [Tomasello, 2019]. In addition, they might be essential for human children already early on due to the social system they grow up in. Other than in most mammals, human infants are born to mothers that are not always fully committed to them and are depending on the care of additional other individuals. It is therefore crucial to engage potential caregivers to interact with them. The better an infant is at engaging caregivers, the more care it receives and the higher its chances of survival are. Human infants are therefore very attentive to others and quick to learn which behaviours are successfully gaining the attention of others [Hrdy, 2009; Hrdy and Burkart, *in rev.*].

The stark contrast in cooperative behaviour between humans and chimpanzees is evident in the fact that chimpanzees rarely share food actively. Mothers allow their infants to take some scraps of their food, but these usually consist of cheap parts like husks or other leftovers. Male chimpanzees do hunt collectively, but the prey will not be shared between all the hunters, but kept by the one individual that did the kill. Allies can get some of the carcasses by tolerated theft and scrounging, but the possessor of the meat almost never hands over parts voluntarily, but mainly due to the harassment of the begging individuals [Gilby, 2006; Jaeggi et al., 2010; Jaeggi and van Schaik, 2011]. Also, chimpanzees do not share information, even if they are proficient in a language system, like the language-trained apes. A study about the communicative content of different language trained apes found that these animals mainly produced requests towards their human communication partners, while the rest of the communication was usually naming objects [Rivas, 2005]. Also in a group of chimpanzees all proficient in sign language, the animals rarely signed towards each other, and if they did, they also mainly produced requests (to play or about food) [Fouts et al., 1984]. One exception seems to be the warning calls chimpanzees provide to group mates when those are unaware of danger (a snake in this study). In this case, it seems like the chimpanzees did intentionally signal the danger, and adjusted the signalling to the state of knowledge of the recipient [Crockford et al., 2012; Crockford et al., 2015; Schel et al., 2013]. Still, this seems to be rather the exception than the rule. Chimpanzees do not systematically provide information to their offspring, even if those would benefit from such information when learning complex foraging strategies and techniques. Even though some studies report teaching behaviour in chimpanzees, this is usually limited to effects like letting infants interact with the tools or other artefacts of the behaviour of the mothers, and never consists of demonstrating the behaviour or instructing the infant [Boesch et al., 2019], which makes it markedly different from teaching in humans. In general, it is fair to say that chimpanzees are not a cooperative species. This is not to say that chimpanzees are unable to cooperate or never do cooperate spontaneously

(see [Gibson, 2012] for some example of chimpanzee cooperation). They are very well capable of doing so, but they are hardly motivated to do so, and almost never cooperate just for the benefit of others.

Quite differently from chimpanzees, marmosets and tamarins cooperate regularly and quite easily. This is reflected most fundamentally in their social organization as cooperative breeders, where some adult individuals forgo reproduction and help others (often their parents) to raise their offspring by caring for those. Help in infant care includes carrying and cleaning the dependent offspring, sharing food with them and offer protection to the whole group by vigilance and territory defence behaviour [Snowdon and Ziegler, 2007]. The transfer of the babies from one caregiver to another has to be highly coordinated, and both animals have to cooperate in the task, as failing could end in losing the offspring and would be highly costly. To be successful, two different preconditions have to be met: (1) The current carrier might signal the need or want to hand over the infants while a potential new carrier perceives the signal and reacts accordingly. Animals therefore have to monitor their group members to perceive their needs as well as show a certain willingness to fulfil those when possible. (2) Further, both animals need a high tolerance for close proximity that is required to hand the infants over (usually, the new carrier plucks them directly off the back of the current carrier) [Burkart and van Schaik, 2010]. This close monitoring of group members might actually be the cause that facilitate social learning in marmosets discussed above [Burkart et al., 2009]. Generally, marmosets show high performance in cooperative tasks that require a socio-cognitive mind set allowing for close inspection and contact of and with group members. Both in a board pulling task as well as in the so-called group service task, where an individual can provide food to its group members without having access to the food itself, marmosets were highly successful in cooperating and provided group members with food regularly. Other than cooperatively breeding callitrichids, only human children showed a similar tendency of cooperating, while non-cooperatively breeding primates, including the apes, did not show any such tendencies [Burkart et al., 2014]. The cooperative nature of marmosets also becomes evident in the high readiness to share food with other (mainly immature) group members. As discussed above, in most primate species active food sharing is rare outside the context of mother-infant dyads and mainly consists of tolerated theft or food transfer due to harassment, rather than active provisioning [Jaeggi and van Schaik, 2011]. Callitrichids on the other hand share high quality food regularly and proactively, and most members of the group share food with infants, with a peak provisioning phase at the time of weaning [Guerreiro et al., 2019; Jaeggi and Gurven, 2013]. Marmoset helpers seem genuinely concerned about immatures, and increased their care effort in a situation when no other caregiver was available [Brügger et al., 2018]. Also, callitrichids are the only primates that have a specific vocalization for both food offering as well as for begging [Jaeggi and Gurven, 2013]. Adult group members that are willing to share food produce specific vocalizations that attract immatures and are a signal that the immature is allowed to take the food parcel the adult is offering, usually by holding it in its hand [Brown et al., 2004]. It even seems that sharing in callitrichids is not limited to food sharing, but that they also share information with their infants [Burkart et al., 2018]. In golden lion tamarins, food offering calls seem to be even involved in a very specific form of teaching. The adults first use calls to attract infants to share already captured prey, but as infants get older, the adult call the infants to draw their attention to specific feeding substrates, where the infant then catches the prey localized by the adult beforehand [Rapaport, 2011]. Another example comes from cotton top tamarins, which provide their group members with information about the quality of food. They produce a very distinct visual signal when sampling unpalatable food, and sometimes even produced warning calls, which can be used by group members to learn about the food quality. As a consequence, cotton top tamarins are one of the only primate species shown to avoid unpalatable food not only by sampling it themselves, but also by observing others doing

so, respectively by observing their disgust reaction [Snowdon and Boe, 2003]. Opposite to chimpanzees but similar to humans, marmosets are highly cooperative in their very nature.

5.4.2 Language space in different species

Summing up and bringing together the three previous paragraphs, we see the following picture: For a communication system like human language to emerge, three conditions have to be met: Content that is worth being shared, cognition that allows both the understanding of context as well as the management of a complex communication system, and a cooperative mind-set that facilitates sharing. While humans have a material and social culture that provides a lot of content, have high cognitive abilities and are highly cooperative, chimpanzees do have some potential content as well as cognitive skills required for language but show little cooperative tendencies. Marmosets on the other hand have less elaborated content and are rather limited in their cognitive abilities, but are highly cooperative. Table 5.1 provides an overview over how each of these three conditions is met in humans, chimpanzees and common marmosets.

Table 5.1 Overview over the cognitive potential, communicative content as well as cooperative attitude of humans, chimpanzees and marmosets discussed in the previous sections.

	Humans	**Chimpanzees**	**Marmosets**
Cognition			
Physical domain	High cognitive performance	Cognitive skills comparable to 2 years old human children	Rather low, corresponding to their brain size
Social domain	High cognitive performance, starting at young age	Low, do not reach the levels of 2 years old human children	High, usually out-compete larger-brained sister taxa in task testing the social domain
Content			
Culture	Extended cumulative culture	Culture established	Not reported
Foraging	Complicated foraging mechanism	Extractive foraging	Some species with some extractive foraging, but usually simple foraging strategies
Tool use	Extended use and production of tools	Some production and regular use of tools	Not systematic tool use reported so far
Attitude			
Cooperation	Highly cooperative	Not especially cooperative, although cooperative behaviour can occur	Highly cooperative
Teaching	Even though occurring in diverse forms, teaching is a human universal, children start teaching at young age	No teaching	Some teaching (or at least information donation)

Only humans are highly proficient in all the conditions that allow a complex communication system to be positively selected for: High cognitive abilities like in all great apes are combined with the cooperative social system that allows positive attention and concern towards others, which can ultimately lead to the sharing of information with other individuals (see also [Burkart et al., 2018; Burkart et al., 2009]).

Even though chimpanzees seem to share with humans the cognitive capacities as well as the communicative flexibility necessary to acquire language (irrelevant whether vocal or gestural) as well as some content (mainly foraging techniques) that might be beneficial if they could share it more easily, chimpanzees lack the basic psychology that allows them to use language in a cooperative way. Or, to put it in another way, it seems like chimpanzees have the necessary conditions for language, but lack the cooperative mind-set that allows them apply them.
Marmosets on the other hand do share the cooperative mind-set with humans, which most likely also explains their – for nonhuman primate standards – enhanced vocal communication and flexibility as well as their high level of information donation. They are highly motivated to pay attention to others and share with group members in a variety of contexts. Still, as both content as well as cognitive capacities are relatively restricted in common marmosets, the communication system – although more elaborate than in most other nonhuman primates – remains fairly simple compared to language, but is sufficient for the needs of common marmosets. Only where each of this three conditions overlap to a large part, they create a space within which language can evolve (Figure 5.1).

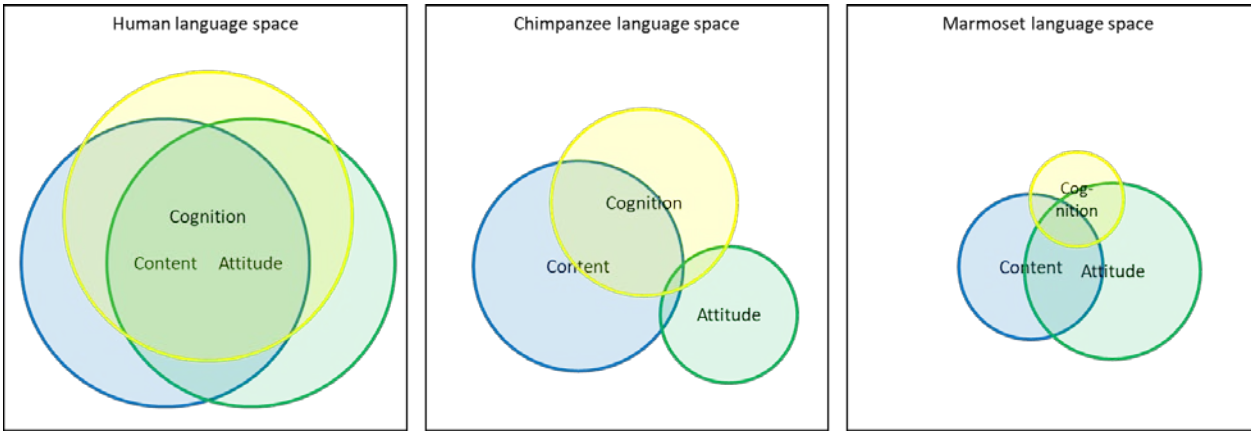

Figure 5.1: The three elements of cognition, communicative content and communication attitude are highly linked. While in chimpanzees content is mainly generated due to their high cognitive abilities, both the cognitive abilities and communicative content in common marmosets are strongly based on their cooperative attitude. In humans, all three elements overlap and interact strongly. Only in case of such a strong overlap, all the necessary preconditions were met to allow for language evolution. While chimpanzees are cognitively skilled and have potentially interesting content, they lack the communicative attitude that would allow sharing information. Marmoset on the other hand have a very cooperative attitude, but are restricted in their cognitive skills and potential contexts.

5.5 Conclusions

Primate vocal communication, even though fundamentally different from human language, shows a series of astounding similarities with the latter. By studying comparable behaviours and potential precursors of typical language behaviour in other species, we can try to deduce the most likely scenario for language evolution. Many preconditions for language evolution are already present in other primate species, like a basic predisposition to understand grammar and syntax, the understanding of arbitrary signals, sharing information with group members as well as a predisposition to vocal learning. Primates follow a variety of conversational rules, combine call signals, show vocal flexibility in different areas and many of these behaviours seem to be at least partially learned during development and influenced by social input.

Vocal learning is a very important part of human language and language acquisition. The current opinion about vocal learning in primates still is that primate vocalizations are largely innate and fixed. According to an increasing body of literature as well as to the results from my studies here, I would argue that it is time to challenge this view. Even though primates do not seem to learn new vocalizations, at least some species are able to modify their vocalizations due to social inputs. Their calls are therefore neither as inflexible nor genetically fixed as commonly assumed, and the results summarized in this thesis show that the foundation of vocal learning might already have been present in the common ancestor of human and nonhuman primates. Further, many primates show conversational rules in communication (avoiding overlapping calls, taking turns, modifying communication due to social influences), and there is evidence that at least certain primate species have volitional control over their vocal output [Tallerman and Gibson, 2012], indicating that vocal communication in primates is flexible to some extent in different frameworks.

The transition from those different pre-language elements to a fully-fledged language systems somewhere during human evolution was most likely rendered possible by the combination of the following elements: an ape-like cognition that provided the capacity to both manage a language system as well as to produce behaviours that lead to content worth sharing, combined with the increased social cognitive abilities, promoting positive attention and proximity towards others, that originated in the cooperative breeding system of early hominins. The combination of content, cognition and communicative attitude made sharing of information possible and beneficial, and could therefore create one of the most unique human behaviours.

References

Baker MC, Cunningham MA. 1985. The Biology of Bird-Song Dialects. Behavioral and Brain Sciences 8(1):85-100.

Barton K. 2009. Mumin: Multi-Model Inference, R Package Version 0.12. 0. http://r-forge r-project org/projects/mumin/.

Berger CR, Calabrese RJ. 1974. Some Explorations in Initial Interaction and Beyond: Toward a Developmental Theory of Interpersonal Communication. Human Communication Research 1(2):99-112.

Berwick RC, Okanoya K, Beckers GJ, Bolhuis JJ. 2011. Songs to Syntax: The Linguistics of Birdsong. Trends in Cognitive Sciences 15(3):113-121.

Bezerra BM, Souto A. 2008. Structure and Usage of the Vocal Repertoire of *Callithrix Jacchus*. International Journal of Primatology 29(3):671-701.

Bockelman WY, Schilling D. 1984. Inheritance of Stereotyped Gibbon Calls. Nature 312:634-636.

Boersma P, Weenink D. 2009. Praat: Doing Phonetics by Computer (Version 5.3.75)[Computer Program]. Retrieved April 30, 2014 from http://www.praat.org.

Boesch C, Bombjaková D, Meier A, Mundry R. 2019. Learning Curves and Teaching When Acquiring Nut-Cracking in Humans and Chimpanzees. Scientific Reports 9.

Bouchet H, Koda H, Lemasson A. 2017. Age-Dependent Change in Attention Paid to Vocal Exchange Rules in Japanese Macaques. Animal Behaviour 129:81-92.

Boughman JW. 1997. Greater Spear-Nosed Bats Give Group-Distinctive Calls. Behavioral Ecology and Sociobiology 40(1):61-70.

Boughman JW, Wilkinson GS. 1998. Greater Spear-Nosed Bats Discriminate Group Mates by Vocalizations. Animal Behaviour 55(6):1717-1732.

Boyette AH, Hewlett BS. 2018. Teaching in Hunter-Gatherers. Review of Philosophy and Psychology 9(4):771-797.

Briefer EF, McElligot AG. 2011. Indicators of Age, Body Size and Sex in Goat Kid Calls Revealed Using the Source–Filter Theory. Applied Animal Behaviour Science 133(3):175-185.

Brown GR, Almond RE, Bergen YV. 2004. Begging, Stealing, and Offering: Food Transfer in Nonhuman Primates. Advances in the Study of Behavior 34(265):e295.

Brügger R, Kappeler-Schmalzriedt T, Burkart J. 2018. Reverse Audience Effects on Helping in Cooperatively Breeding Marmoset Monkeys. Biology Letters 14(3):20180030.

Brumm H. 2004. Acoustic Communication in Noise: Regulation of Call Characteristics in a New World Monkey. Journal of Experimental Biology 207(3):443-448.

Burkart J, Guerreiro Martins E, Miss F, Zuercher Y. 2018. From Sharing Food to Sharing Information. Cooperative Breeding and Language Evolution. Interaction Studies 19(1-2).

Burkart J, Heschl A. 2006. Geometrical Gaze Following in Common Marmosets (*Callithrix Jacchus*). Journal of Comparative Psychology 120(2):120.

Burkart J, Kupferberg A, Glasauer S, van Schaik C. 2012. Even Simple Forms of Social Learning Rely on Intention Attribution in Marmoset Monkeys (*Callithrix Jacchus*). Journal of Comparative Psychology 126(2):129.

Burkart JM, Allon O, Amici F, Fichtel C, Finkenwirth C, Heschl A, Huber J, Isler K, Kosonen Z, Martins E, Meulman EJ, Richiger R, Rueth K, Spillmann B, Wiesendanger S, van Schaik CP. 2014. The Evolutionary Origin of Human Hyper-Cooperation. Nature Communications 5:4747.

Burkart JM, Finkenwirth C. 2015. Marmosets as Model Species in Neuroscience and Evolutionary Anthropology. Neuroscience Research 93:8-19.

Burkart JM, Heschl A. 2007. Understanding Visual Access in Common Marmosets, *Callithrix Jacchus*: Perspective Taking or Behaviour Reading? Animal Behaviour 73(3):457-469.

Burkart JM, Hrdy SB, van Schaik CP. 2009. Cooperative Breeding and Human Cognitive Evolution. Evolutionary Anthropology: Issues, News, and Reviews 18(5):175-186.

Burkart JM, van Schaik CP. 2010. Cognitive Consequences of Cooperative Breeding in Primates? Animal Cognition 13(1):1-19.

Candiotti A, Zuberbühler K, Lemasson A. 2012. Convergence and Divergence in Diana Monkey Vocalizations. Biology Letters 8(3):382-5.

Capitanio JP, Blozis SA, Snarr J, Steward A, McCowan BJ. 2017. Do "Birds of a Feather Flock Together" or Do "Opposites Attract"? Behavioral Responses and Temperament Predict Success in Pairings of Rhesus Monkeys in a Laboratory Setting. American Journal of Primatology: Official Journal of the American Society of Primatologists 79(1):1-11.

Chen HC, Kaplan G, Rogers L. 2009. Contact Calls of Common Marmosets (*Callithrix Jacchus*): Influence of Age of Caller on Antiphonal Calling and Other Vocal Responses. American Journal of Primatology: Official Journal of the American Society of Primatologists 71(2):165-170.

Choi JY, Takahashi DY, Ghazanfar AA. 2015. Cooperative Vocal Control in Marmoset Monkeys Via Vocal Feedback. Journal of Neurophysiology 114(1):274-83.

Chow CP, Mitchell JF, Miller CT. 2015. Vocal Turn-Taking in a Non-Human Primate Is Learned During Ontogeny. Proceedings Biological sciences / The Royal Society 282(1807):20150069.

Christiansen MH, Kirby S. 2003. Language Evolution: The Hardest Problem in Science? Studies in the Evolution of Language 3:1-15.

Clarke E, Reichard UH, Zuberbühler K. 2006. The Syntax and Meaning of Wild Gibbon Songs. PloS one 1(1):e73.

Crockford C, Boesch C. 2005. Call Combinations in Wild Chimpanzees. Behaviour:397-421.

Crockford C, Herbinger I, Vigilant L, Boesch C. 2004. Wild Chimpanzees Produce Group-Specific Calls: A Case for Vocal Learning? Ethology 110:221—243.

Crockford C, Wittig RM, Mundry R, Zuberbühler K. 2012. Wild Chimpanzees Inform Ignorant Group Members of Danger. Current Biology 22(2):142-146.

Crockford C, Wittig RM, Zuberbühler K. 2015. An Intentional Vocalization Draws Others' Attention: A Playback Experiment with Wild Chimpanzees. Animal Cognition 18(3):581-591.

Csibra G, Gergely G. 2011. Natural Pedagogy as Evolutionary Adaptation. Philosophical Transactions of the Royal Society B: Biological Sciences 366(1567):1149-1157.

de la Torre S, Snowdon CT. 2002. Environmental Correlates of Vocal Communication of Wild Pygmy Marmosets, *Cebuella Pygmaea*. Animal Behaviour 63(5):847-856.

de la Torre S, Snowdon CT. 2009. Dialects in Pygmy Marmosets? Population Variation in Call Structure. American Journal of Primatology 71(4):333-42.

Deaner RO, Isler K, Burkart J, van Schaik C. 2007. Overall Brain Size, and Not Encephalization Quotient, Best Predicts Cognitive Ability across Non-Human Primates. Brain, Behavior and Evolution 70(2):115-124.

Deaner RO, van Schaik CP, Johnson V. 2006. Do Some Taxa Have Better Domain-General Cognition Than Others? A Meta-Analysis of Nonhuman Primate Studies. Evolutionary Psychology 4(1):147470490600400114.

Delgado RA. 2007. Geographic Variation in the Long Calls of Male Orangutans (Pongo Spp.). Ethology 113(5):487-498.

Demartsev V, Strandburg-Peshkin A, Ruffner M, Manser M. 2018. Vocal Turn-Taking in Meerkat Group Calling Sessions. Current Biology 28(22):3661-3666. e3.

Díaz-Muñoz SL, Bales KL. 2016. "Monogamy" in Primates: Variability, Trends, and Synthesis: Introduction to Special Issue on Primate Monogamy. American Journal of Primatology: Official Journal of the American Society of Primatologists 78(3):283-287.

Dunbar RI. 2003. The Origin and Subsequent Evolution of Language. Studies in the Evolution of Language 3:219-234.

Dyble M, Thompson J, Smith D, Salali GD, Chaudhary N, Page AE, Vinicuis L, Mace R, Migliano AB. 2016. Networks of Food Sharing Reveal the Functional Significance of Multilevel Sociality in Two Hunter-Gatherer Groups. Current Biology 26(15):2017-2021.

Egnor SE, Hauser MD. 2004. A Paradox in the Evolution of Primate Vocal Learning. TRENDS in Neurosciences 27(11):649-54.

Eliades SJ, Miller CT. 2017. Marmoset Vocal Communication: Behavior and Neurobiology. Developmental Neurobiology 77(3):286-299.

Eliades SJ, Wang X. 2012. Neural Correlates of the Lombard Effect in Primate Auditory Cortex. The Journal of Neuroscience : the Official Journal of the Society for Neuroscience 32(31):10737-48.

Elowson AM, Tannenbaum PL, Snowdon CT. 1991. Food-Associated Calls Correlate with Food Preferences in Cotton-Top Tamarins. Animal Behaviour 42(6):931-937.

Elowson MA, Snowdon CT. 1994. Pygmy Marmosets, *Cebuella Pygmaea*, Modify Vocal Structure in Response to Changed Social Environment. Animal Behaviour 47(6):1267-1277.

Elowson MA, Snowdon CT, Lazaro-Perea C. 1998a. 'Babbling' and Social Context in Infant Monkeys: Parallels to Human Infants. Trends in Cognitive Sciences 2(1):31 -37.

Elowson MA, Snowdon CT, Lazaro-Perea C. 1998b. Infant 'Babbling' in a Nonhuman Primate: Complex Vocal Sequences with Repeated Call Types. Behaviour 135:643-664.

Engesser S, Holub JL, O'Neill LG, Russell AF, Townsend SW. 2019. Chestnut-Crowned Babbler Calls Are Composed of Meaningless Shared Building Blocks. Proceedings of the National Academy of Sciences:201819513.

Engesser S, Ridley AR, Townsend SW. 2016. Meaningful Call Combinations and Compositional Processing in the Southern Pied Babbler. Proceedings of the National Academy of Sciences 113(21):5976-5981.

Ey E, Rahn C, Hammerschmidt K, Fischer J. 2009. Wild Female Olive Baboons Adapt Their Grunt Vocalizations to Environmental Conditions. Ethology 115(5):493-503.

Fedurek P, Slocombe KE. 2011. Primate Vocal Communication: A Useful Tool for Understanding Human Speech and Language Evolution? Human Biology 83(2):153-173.

Filatova OA, Fedutin ID, Burdin AM, Hoyt E. 2007. The Structure of the Discrete Call Repertoire of Killer Whales *Orcinus Orca* from Southeast Kamchatka. Bioacoustics 16(3):261-280.

Finkenwirth C, Burkart JM. 2018. Why Help? Relationship Quality, Not Strategic Grooming Predicts Infant-Care in Group-Living Marmosets. Physiology & Behavior 193:108-116.

Fischer J, Wheeler BC, Higham JP. 2015. Is There Any Evidence for Vocal Learning in Chimpanzee Food Calls? Current Biology 25(21):R1028-R1029.

Fitch TW. 2000. The Evolution of Speech: A Comparative Review. Trends in Cognitive Sciences 4(7):258–267.

Fitch WT. 2005. The Evolution of Language: A Comparative Review. Biology and Philosophy 20(2-3):193-203.

Fitch WT. 2007. Evolving Meaning: The Roles of Kin Selection, Allomothering and Paternal Care in Language Evolution. Emergence of Communication and Language: Springer. p 29-51.

Ford JK. 1991. Vocal Traditions among Resident Killer Whales (*Orcinus Orca*) in Coastal Waters of British Columbia. Canadian Journal of Zoology 69(6):1454-1483.

Fouts RS, Fouts DH, Schoenfeld D. 1984. Sign Language Conversational Interaction between Chimpanzees. Sign Language Studies 42(1):1-12.

Geissmann T. 1984. Inheritance of Song Parameters in the Gibbon Song, Analysed in 2 Hybrid Gibbons (*Hylobates Pileatus× H. Lar*). Folia Primatologica 42(3-4):216-235.

Gibson KR. 2012. Language or Protolanguage? A Review of the Ape Language Literature. The Oxford Handbook of Language Evolution.

Gilby IC. 2006. Meat Sharing among the Gombe Chimpanzees: Harassment and Reciprocal Exchange. Animal Behaviour 71(4):953-963.

Guerreiro EM, Moura A, Finkenwirth C, Griesser M, Burkart J. 2019. Food Sharing Patterns in Three Species of Callitrichid Monkeys (*Callithrix Jacchus, Leontopithecus Chrysomelas, Saguinus Midas)*: Individual and Species Differences. Journal of Comparative Psychology (Washington, DC: 1983).

Gultekin YB, Hage SR. 2017. Limiting Parental Feedback Disrupts Vocal Development in Marmoset Monkeys. Nature Communications 8:14046.

Gultekin YB, Hage SR. 2018. Limiting Parental Interaction During Vocal Development Affects Acoustic Call Structure in Marmoset Monkeys. Science Advances 4(4):eaar4012.

Hammerschmidt K, Freudenstein T, Jürgens U. 2001. Vocal Development in Squirrel Monkeys. Behaviour 138(9):1179-1204.

Hammerschmidt K, Newman JD, Champoux M, Suomi SJ. 2000. Changes in Rhesus Macaque 'Coo' Vocalizations During Early Development. Ethology 106(10):873-886.

Haun D, Over H. 2015. Like Me: A Homophily-Based Account of Human Culture. Epistemological Dimensions of Evolutionary Psychology: Springer. p 117-130.

Hauser MD, Chomsky N, Fitch WT. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? Science 298(5598):1569-1579.

Hedwig D, Mundry R, Robbins MM, Boesch C. 2015. Contextual Correlates of Syntactic Variation in Mountain and Western Gorilla Close-Distance Vocalizations: Indications for Lexical or Phonological Syntax? Animal Cognition 18(2):423-435.

Heldstab SA, Isler K, Burkart JM, van Schaik CP. 2019. Allomaternal Care, Brains and Fertility in Mammals: Who Cares Matters. Behavioral Ecology and Sociobiology 73(6):71.

Henry L, Barbu S, Lemasson A, Hausberger M. 2015. Dialects in Animals: Evidence, Development and Potential Functions. Animal Behavior and Cognition 2(2):132--155.

Herrmann E, Call J, Hernàndez-Lloreda MV, Hare B, Tomasello M. 2007. Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis. Science 317(5843):1360-1366.

Hotchkin C, Parks S. 2013. The Lombard Effect and Other Noise-Induced Vocal Modifications: Insight from Mammalian Communication Systems. Biological Reviews 88(4):809-824.

Hotchkin CF, Parks SE, Weiss DJ. 2015. Noise-Induced Frequency Modifications of Tamarin Vocalizations: Implications for Noise Compensation in Nonhuman Primates. PLoS One 10(6):e0130211.

Hrdy SB. 2009. Mothers and Others: Harvard University Press.

Hrdy SB, Burkart JM. *in rev.* The Emergence of Emotionally Modern Humans: Implications for Language and Learning. Philosophical Transactions of the Royal Society B: Biological Sciences.

Isler K, van Schaik CP. 2014. How Humans Evolved Large Brains: Comparative Evidence. Evolutionary Anthropology: Issues, News, and Reviews 23(2):65-75.

Jaeggi AV, Gurven M. 2013. Natural Cooperators: Food Sharing in Humans and Other Primates. Evolutionary Anthropology: Issues, News, and Reviews 22(4):186-195.

Jaeggi AV, Stevens JM, van Schaik CP. 2010. Tolerant Food Sharing and Reciprocity Is Precluded by Despotism among Bonobos but Not Chimpanzees. American Journal of Physical Anthropology 143(1):41-51.

Jaeggi AV, van Schaik CP. 2011. The Evolution of Food Sharing in Primates. Behavioral Ecology and Sociobiology 65(11):2125.

Janik VM. 2012. Vocal Communication and Cognition in Cetaceans. The Oxford Handbook of Language Evolution.

Janik VM, Slater PJ. 1997. Vocal Learning in Mammals. Advances in the Study of Behaviour 26:59-100.

Janik VM, Slater PJ. 2000. The Different Roles of Social Learning in Vocal Communication. Animal Behaviour 60(1):1-11.

Jarvis ED. 2006. Selection for and against Vocal Learning in Birds and Mammals. Ornithological Science 5(1):5-14.

Jenkins PF. 1978. Cultural Transmission of Song Patterns and Dialect Development in a Free-Living Bird Population. Animal Behaviour 26:50-78.

Jones BS, Harris DH, Catchpole CK. 1993. The Stability of the Vocal Signature in Phee Calls of the Common Marmoset, *Callithrix Jacchus*. American Journal of Primatology 31(1):67-75.

Kako E. 1999. Elements of Syntax in the Systems of Three Language-Trained Animals. Animal Learning & Behavior 27(1):1-14.

Kato M, Okanoya K, Koike T, Sasaki E, Okano H, Watanabe S, Iriki A. 2014. Human Speech-and Reading-Related Genes Display Partially Overlapping Expression Patterns in the Marmoset Brain. Brain and Language 133:26-38.

Katsu N, Yamada K, Okanoya K, Nakamichi M. 2018. Temporal Adjustment of Short Calls According to a Partner During Vocal Turn-Taking in Japanese Macaques. Current Zoology 65(1):99-105.

Kershenbaum A, Ilany A, Blaustein L, Geffen E. 2012. Syntactic Structure and Geographical Dialects in the Songs of Male Rock Hyraxes. Proceedings of the Royal Society B: Biological Sciences 279(1740):2974-2981.

Kitzmann CD, Caine NG. 2009. Marmoset (*Callithrix Geoffroyi*) Food-Associated Calls Are Functionally Referential. Ethology 115(5):439-448.

Koda H, Lemasson A, Oyakawa C, Rizaldi, Pamungkas J, Masataka N. 2013. Possible Role of Mother-Daughter Vocal Interactions on the Development of Species-Specific Song in Gibbons. PLoS One 8(8):e71432.

Koops K, Wrangham RW, Cumberlidge N, Fitzgerald MA, van Leeuwen KL, Rothman JM, Matsuzawa T. 2019. Crab-Fishing by Chimpanzees in the Nimba Mountains, Guinea. Journal of Human Evolution.

Koski SE, Burkart JM. 2015. Common Marmosets Show Social Plasticity and Group-Level Similarity in Personality. Scientific Reports 5:8878.

Kuhl PK. 2004. Early Language Acquisition: Cracking the Speech Code. Nature Reviews Neuroscience 5(11):831-843.

Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. Lmertest Package: Tests in Linear Mixed Effects Models. 2017 82(13):26.

Laland KN, Hoppitt W. 2003. Do Animals Have Culture? Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews 12(3):150-159.

Lameira AR, Delgado RA, Wich SA. 2010. Review of Geographic Variation in Terrestrial Mammalian Acoustic Signals: Human Speech Variation in a Comparative Perspective. Journal of Evolutionary Psychology 8(4):309-332.

Laporte MN, Zuberbühler K. 2010. Vocal Greeting Behaviour in Wild Chimpanzee Females. Animal Behaviour 80(3):467-473.

Laporte MN, Zuberbühler K. 2011. The Development of a Greeting Signal in Wild Chimpanzees. Developmental Science 14(5):1220-1234.

Lazaro-Perea C. 2001. Intergroup Interactions in Wild Common Marmosets, *Callithrix Jacchus*: Territorial Defence and Assessment of Neighbours. Animal Behaviour 62(1):11-21.

Le Boeuf BJ, Petrinovich LF. 1974. Dialects of Northern Elephant Seals, *Mirounga Angustirostris:* Origin and Reliability. Animal Behaviour 22(3):656-663.

Lemasson A, Gandon E, Hausberger M. 2010. Attention to Elders' Voice in Non-Human Primates. Biology Letters 6(3):328.

Lemasson A, Glas L, Barbu S, Lacroix A, Guilloux M, Remeuf K, Koda H. 2011a. Youngsters Do Not Pay Attention to Conversational Rules: Is This So for Nonhuman Primates? Scientific Reports 1:22.

Lemasson A, Hausberger M. 2004. Patterns of Vocal Sharing and Social Dynamics in a Captive Group of Campbell's Monkeys (*Cercopithecus Campbelli Campbelli*). Journal of Comparative Psychology 118(3):347-59.

Lemasson A, Jubin R, Masataka N, Arlet M. 2016. Copying Hierarchical Leaders' Voices? Acoustic Plasticity in Female Japanese Macaques. Scientific Reports 6:21289.

Lemasson A, Ouattara K, Petit EJ, Zuberbuhler K. 2011b. Social Learning of Vocal Structure in a Nonhuman Primate? BMC Evolutionary Biology 11:362.

Liao DA, Zhang YS, Cai LX, Ghazanfar AA. 2018. Internal States and Extrinsic Factors Both Determine Monkey Vocal Production. Proceedings of the National Academy of Sciences:201722426.

Liebal K, Waller BM, Slocombe KE, Burrows AM. 2014. Primate Communication: A Multimodal Approach: Cambridge University Press.

Lieblich AK, Symmes D, Newman JD, Shapiro M. 1980. Development of the Isolation Peep in Laboratory-Bred Squirrel Monkeys. Animal Behaviour 28:1 - 9.

Marshall-Pescini S, Dale R, Quervel-Chaumette M, Range F. 2016. Critical Issues in Experimental Studies of Prosociality in Non-Human Species. Animal Cognition 19(4):679-705.

Marx V. 2016. Neurobiology: Learning from Marmosets. Nature Publishing Group.

Masataka N, Biben M, Symmes D. 1986. Temporal and Structural Analysis of Affiliative Vocal Exchanges in Squirrel Monkeys (*Saimiri Sciureus*). Behaviour 98(1-4):259-273.

Massen JJM, Koski SE. 2014. Chimps of a Feather Sit Together: Chimpanzee Friendships Are Based on Homophily in Personality. Evolution and Human Behavior 35(1):1-8.

Miller CT, Freiwald WA, Leopold DA, Mitchell JF, Silva AC, Wang X. 2016. Marmosets: A Neuroscientific Model of Human Social Behavior. Neuron 90(2):219-33.

Miller CT, Mandel K, Wang X. 2010. The Communicative Content of the Common Marmoset Phee Call During Antiphonal Calling. American Journal of Primatology: Official Journal of the American Society of Primatologists 72(11):974-80.

Mitani JC, Gros-Louis J. 1998. Chorusing and Call Convergence in Chimpanzees: Tests of Three Hypotheses. Behaviour 135(8):1041-1064.

Mitani JC, Hasegawa T, Gros-Louis J, Marler P, Byrne R. 1992. Dialects in Wild Chimpanzees? American Journal of Primatology: Official Journal of the American Society of Primatologists 27:233-243.

Mitani JC, Hunley K, Murdoch M. 1999. Geographic Variation in the Calls of Wild Chimpanzees: A Reassessment. American Journal of Primatology: Official Journal of the American Society of Primatologists 47(2):133-151.

Mundry R, Sommer C. 2007. Discriminant Function Analysis with Nonindependent Data: Consequences and an Alternative. Animal Behaviour 74(4):965-976.

Nakagawa S, Schielzeth H. 2013. A General and Simple Method for Obtaining R2 from Generalized Linear Mixed-Effects Models. Methods in Ecology and Evolution 4(2):133-142.

Nelson DGK, Hirsh-Pasek K, Jusczyk PW, Cassidy KW. 1989. How the Prosodic Cues in Motherese Might Assist Language Learning. Journal of Child Language 16(1):55-68.

Noad MJ, Cato DH, Bryden M, Jenner M-N, Jenner KCS. 2000. Cultural Revolution in Whale Songs. Nature 408(6812):537.

Noë R, de Waal FB, van Hooff JA. 1980. Types of Dominance in a Chimpanzee Colony. Folia Primatologica 34(1-2):90-110.

Norcross J, Newman J. 1997. Social Context Affects Phee Call Production by Nonreproductive Common Marmosets (*Callithrix Jacchus*). American Journal of Primatology: Official Journal of the American Society of Primatologists 43(2):135-146.

Norcross JL, Newmann JD, Cofrancesco lM. 1999. Context and Sex Differences Exist in the Acoustic Structure of Phee Calls by Newly–Paired Common Marmosets (*Callithrix Jacchus*). American Journal of Primatology: Official Journal of the American Society of Primatologists 49:165–181.

Owen C, Rendell L, Constantine R, Noad MJ, Allen J, Andrews O, Garrigue C, Michael Poole M, Donnelly D, Hauser N. 2019. Migratory Convergence Facilitates Cultural Transmission of Humpback Whale Song. Royal Society Open Science 6(9):190337.

Owren MJ, Dieter JA, Seyfarth RM, Cheney DL. 1992. 'Food' Calls Produced by Adult Female Rhesus *(Macaca Mulatta)* and Japanese *(M. Fuscata)* Macaques, Their Normally-Raised Offspring and Offspring Cross-Fostered between Species. Behaviour 120(3-4):218-231.

Parkinson C, Kleinbaum AM, Wheatley T. 2018. Similar Neural Responses Predict Friendship. Nature Communications 9(1):332.

Pearl D, Fenton M. 1996. Can Echolocation Calls Provide Information About Group Identity in the Little Brown Bat (*Myotis Lucifugus*)? Canadian Journal of Zoology 74(12):2184-2192.

Pepperberg IM. 2012. Evolution of Communication and Language: Insights from Parrots and Songbirds. The Oxford Handbook of Language Evolution.

Pfenning AR, Hara E, Whitney O, Rivas MV, Wang R, Roulhac PL, Howard JT, Wirthlin M, Lovell PV, Ganapathy G. 2014. Convergent Transcriptional Specializations in the Brains of Humans and Song-Learning Birds. Science 346(6215):1256846.

Pika S, Klein H, Bunel S, Baas P, Théleste E, Deschner T. 2019. Wild Chimpanzees (*Pan Troglodytes Troglodytes*) Exploit Tortoises (*Kinixys Erosa*) Via Percussive Technology. Scientific Reports 9(1):7661.

Pistorio AL, Vintch B, Wang X. 2006. Acoustic Analysis of Vocal Development in a New World Primate, the Common Marmoset (*Callithrix Jacchus*). The Journal of the Acoustical Society of America 120(3):1655.

Podos J, Warren PS. 2007. The Evolution of Geographic Variation in Birdsong. Advances in the Study of Behavior 37:403-458.

Prat Y, Azoulay L, Dor R, Yovel Y. 2017. Crowd Vocal Learning Induces Vocal Dialects in Bats: Playback of Conspecifics Shapes Fundamental Frequency Usage by Pups. PLoS Biology 15(10):e2002556.

Raiche G. 2010. Nfactors: An R Package for Parallel Analysis and Non Graphical Solutions to the Cattell Scree Test (R Package Version 2.3. 3)[Computer Software].

Range F, Huber L. 2007. Attention in Common Marmosets: Implications for Social-Learning Experiments. Animal Behaviour 73(6):1033-1041.

Rapaport LG. 2011. Progressive Parenting Behavior in Wild Golden Lion Tamarins. Behavioral Ecology 22(4):745-754.

Reader SM, Hager Y, Laland KN. 2011. The Evolution of Primate General and Cultural Intelligence. Philosophical Transactions of the Royal Society B: Biological Sciences 366(1567):1017-1027.

Reby D, McComb K. 2003. Anatomical Constraints Generate Honesty: Acoustic Cues to Age and Weight in the Roars of Red Deer Stags. Animal Behaviour 65(3):519-530.

Rendell LE, Whitehead H. 2003. Vocal Clans in Sperm Whales (*Physeter Macrocephalus*). Proceedings of the Royal Society of London Series B: Biological Sciences 270(1512):225-231.

Rivas E. 2005. Recent Use of Signs by Chimpanzees (*Pan Troglodytes*) in Interactions with Humans. Journal of Comparative Psychology 119(4):404.

Roberts AI, Roberts SGB. 2017. Convergence and Divergence in Gesture Repertoires as an Adaptive Mechanism for Social Bonding in Primates. Royal Society Open Science 4(11):170181.

Rogers L, Stewart L, Kaplan G. 2018. Food Calls in Common Marmosets, *Callithrix Jacchus,* and Evidence That One Is Functionally Referential. Animals 8(7):99.

Roupe SL, Pistorio A, Wang X. 2003. Vocal Plasticity Induced by Auditory Deprivation in the Common Marmoset, Program No. 627.7 2003 Abstract Viewer and Itinerary Planner. Society for Neuroscience Online.

Roush RS, Snowdon CT. 2001. Food Transfer and Development of Feeding Behavior and Food-Associated Vocalizations in Cotton-Top Tamarins. Ethology 107(5):415-429.

Roy S, Miller CT, Gottsch D, Wang X. 2011. Vocal Control by the Common Marmoset in the Presence of Interfering Noise. The Journal of Experimental Biology 214(Pt 21):3619-29.

Ruch H, Zürcher Y, Burkart JM. 2018. The Function and Mechanism of Vocal Accommodation in Humans and Other Primates. Biological Reviews 93(2):996-1013.

Rukstalis M, Fite JE, French JA. 2003. Social Change Affects Vocal Structure in a Callitrichid Primate (*Callithrix Kuhlii*). Ethology 109:327—340.

Rukstalis M, French JA. 2005. Vocal Buffering of the Stress Response: Exposure to Conspecific Vocalizations Moderates Urinary Cortisol Excretion in Isolated Marmosets. Hormones and Behavior 47(1):1-7.

Russell JL, Braccini S, Buehler N, Kachin MJ, Schapiro SJ, Hopkins WD. 2005. Chimpanzee (*Pan Troglodytes*) Intentional Communication Is Not Contingent Upon Food. Animal Cognition 8(4):263-272.

Salali GD, Chaudhary N, Bouer J, Thompson J, Vinicius L, Migliano AB. 2019. Development of Social Learning and Play in Bayaka Hunter-Gatherers of Congo. Scientific Reports 9(1):11080.

Savage-Rumbaugh ES, Murphy J, Sevcik RA, Brakke KE, Williams SL, Rumbaugh DM, Bates E. 1993. Language Comprehension in Ape and Child. Monographs of the Society for Research in Child Development:i-252.

Scharff C, Petri J. 2011. Evo-Devo, Deep Homology and Foxp2: Implications for the Evolution of Speech and Language. Philosophical Transactions of the Royal Society B: Biological Sciences 366(1574):2124-2140.

Schel AM, Townsend SW, Machanda Z, Zuberbühler K, Slocombe KE. 2013. Chimpanzee Alarm Call Production Meets Key Criteria for Intentionality. PloS one 8(10):e76674.

Schibler F, Manser MB. 2007. The Irrelevance of Individual Discrimination in Meerkat Alarm Calls. Animal Behaviour 74(5):1259-1268.

Schlenker P, Chemla E, Arnold K, Lemasson A, Ouattara K, Keenan S, Stephan C, Ryder R, Zuberbühler K. 2014. Monkey Semantics: Two 'Dialects' of Campbell's Monkey Alarm Calls. Linguistics and Philosophy 37(6):439-501.

Sewall KB, Young AM, Wright TF. 2016. Social Calls Provide Novel Insights into the Evolution of Vocal Learning. Animal Behaviour 120:163-172.

Seyfarth RM, Cheney DL. 2014. The Evolution of Language from Social Cognition. Current Opinion in Neurobiology 28:5-9.

Seyfarth RM, Cheney DL, Bergman TJ. 2005. Primate Social Cognition and the Origins of Language. Trends in Cognitive Sciences 9(6):264-266.

Smolker R, Pepper JW. 1999. Whistle Convergence among Allied Male Bottlenose Dolphins (Delphinidae, *Tursiops Sp.*). Ethology 105(7):595-617.

Snowdon CT. 2001. Social Processes in Communication and Cognition in Callitrichid Monkeys: A Review. Animal Cognition 4(3-4):247-57.

Snowdon CT. 2013. Language Parallels in New World Primates. Animal Models of Speech and Language Disorders: Springer. p 241-261.

Snowdon CT. 2017. Vocal Communication in Family-Living and Pair-Bonded Primates. Primate Hearing and Communication: Springer. p 141-174.

Snowdon CT, Boe CY. 2003. Social Communication About Unpalatable Foods in Tamarins (*Saguinus Oedipus*). Journal of Comparative Psychology 117(2):142.

Snowdon CT, Cleveland J. 1984. "Conversations" among Pygmy Marmosets. American Journal of Primatology: Official Journal of the American Society of Primatologists 7:15 - 20.

Snowdon CT, de la Torre S. 2002. Multiple Environmental Contexts and Communication in Pygmy Marmosets (*Cebuella Pygmaea*). Journal of Comparative Psychology 116(2):182-188.

Snowdon CT, Elowson MA. 1999. Pygmy Marmoset Modify Their Call Structure When Paired. Ethology 105:782-897.

Snowdon CT, Elowson MA. 2001. 'Babbling' in Pygmy Marmosets: Development after Infancy. Behaviour 138(10):1235-1248.

Snowdon CT, Ziegler TE. 2007. Growing up Cooperatively: Family Processes and Infant Care in Marmosets and Tamarins. Journal of Developmental Processes 2(1):40-66.

Strager H. 1995. Pod-Specific Call Repertoires and Compound Calls of Killer Whales, Orcinus Orca Linnaeus, 1758, in the Waters of Northern Norway. Canadian Journal of Zoology 73(6):1037-1047.

Strauss S, Ziv M. 2012. Teaching Is a Natural Cognitive Ability for Humans. Mind, Brain, and Education 6(4):186-196.

Sugiura H. 1989. Matching of Acoustic Features During the Vocal Exchanges of Coo Calls in Japanese Macaques. Animal Behaviour 55:673–687.

Sugiura H. 1993. Temporal and Acoustic Correlates in Vocal Exchange of Coo Calls in Japanese Macaques. Behaviour 124(3-4):207-225.

Suzuki TN, Wheatcroft D, Griesser M. 2016. Experimental Evidence for Compositional Syntax in Bird Calls. Nature Communications 7:10986.

Takahashi DY, Fenley AR, Ghazanfar AA. 2016. Early Development of Turn-Taking with Parents Shapes Vocal Acoustics in Infant Marmoset Monkeys. Philosophical Transactions of the Royal Society B: Biological Sciences 371(1693):20150370.

Takahashi DY, Fenley AR, Teramoto Y, Narayanan DZ, Borjon JI, Holmes P, Ghazanfar AA. 2015. The Developmental Dynamics of Marmoset Monkey Vocal Production. Science 349(6249):734 - 738.

Takahashi DY, Liao DA, Ghazanfar AA. 2017. Vocal Learning Via Social Reinforcement by Infant Marmoset Monkeys. Current Biology 27(12):1844-1852. e6.

Takahashi DY, Narayanan DZ, Ghazanfar AA. 2013. Coupled Oscillator Dynamics of Vocal Turn-Taking in Monkeys. Current Biology 23(21):2162-2168.

Tallerman M, Gibson KR. 2012. Introduction: The Evolution of Language. The Oxford Handbook of Language Evolution.

Tanaka T, Sugiura H, Masataka N. 2006. Sound Transmission in the Habitats of Japanese Macaques and Its Possible Effect on Population Differences in Coo Calls. Behaviour 143(8):993-1012.

ten Cate C, Okanoya K. 2012. Revisiting the Syntactic Abilities of Non-Human Animals: Natural Vocalizations and Artificial Grammar Learning. Philosophical Transactions of the Royal Society B: Biological Sciences 367(1598):1984-1994.

Thomsen HM, Balsby TJ, Dabelsteen T. 2019. The Imitation Dilemma: Can Parrots Maintain Their Vocal Individuality When Imitating Conspecifics? Behaviour 156(5-8):787-814.

Tomasello M. 2010. Origins of Human Communication: MIT press.

Tomasello M. 2019. Becoming Human: A Theory of Ontogeny: Belknap Press.

Tomasello M, Carpenter M, Call J, Behne T, Moll H. 2005. Understanding and Sharing Intentions: The Origins of Cultural Cognition. Behavioral and Brain Sciences 28:675–735.

Tyack PL. 2007. Convergence of Calls as Animals Form Social Bonds, Active Compensation for Noisy Communication Channels, and the Evolution of Vocal Learning in Mammals. Journal of Comparative Psychology 122(3):319-331.

van Schaik CP. 2016. The Primate Origins of Human Nature: John Wiley & Sons.

Vitale A, Zanzoni M, Queyras A, Chiarotti F. 2003. Degree of Social Contact Affects the Emission of Food Calls in the Common Marmoset (*Callithrix Jacchus*). American Journal of Primatology: Official Journal of the American Society of Primatologists 59(1):21-8.

Voelkl B, Huber L. 2000. True Imitation in Marmosets. Animal Behaviour 60(2):195-202.

Watson CF, Buchanan-Smith HM, Caldwell CA. 2014. Call Playback Artificially Generates a Temporary Cultural Style of High Affiliation in Marmosets. Animal Behaviour 93:163-171.

Watson SK, Townsend SW, Schel AM, Wilke C, Wallace EK, Cheng L, West V, Slocombe KE. 2015a. Reply to Fischer Et Al. Current Biology 25(21):R1030-R1031.

Watson SK, Townsend SW, Schel AM, Wilke C, Wallace EK, Cheng L, West V, Slocombe KE. 2015b. Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees. Current Biology 25(4):495-499.

Weinstein TA, Capitanio JP. 2008. Individual Differences in Infant Temperament Predict Social Relationships of Yearling Rhesus Monkeys (*Macaca Mulatta*). Animal Behaviour 76(2):455-465.

Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, Tutin CE, Wrangham RW, Boesch C. 1999. Cultures in Chimpanzees. Nature 399(6737):682.

Wich SA, Schel AM, De Vries H. 2008. Geographic Variation in Thomas Langur (*Presbytis Thomasi*) Loud Calls. American Journal of Primatology: Official Journal of the American Society of Primatologists 70(6):566-574.

Wickham H. 2009. Ggplot2: Elegant Graphics for Data Analysis Springer-Verlag New York. Version.

Winn HE, Thompson TJ, Cummings WC, Hain J, Hudnall J, Hays H, Steiner WW. 1981. Song of the Humpback Whale — Population Comparisons. Behavioral Ecology and Sociobiology 8(1):41-46.

Winter P, Handley P, Detlev P, Ditmar S. 1973. Ontogeny of Squirrel Monkey Calls under Normal Conditions and under Acoustic Isolation. Behaviour 47(3/4):230-239.

Wobber V, Herrmann E, Hare B, Wrangham R, Tomasello M. 2014. Differences in the Early Cognitive Development of Children and Great Apes. Developmental Psychobiology 56(3):547-573.

Wolfram W. 1998. Dialect in Society. In: Coulmas F, editor. The Handbook of Sociolinguistics. Oxford: Blackwell. p 107–126.

Zhao L, Rad BB, Wang X. 2019. Long-Lasting Vocal Plasticity in Adult Marmoset Monkeys. Proceedings of the Royal Society B 286(1905):20190817.

Zollinger SA, Brumm H. 2011a. The Evolution of the Lombard Effect: 100 Years of Psychoacoustic Research. Behaviour 148(11-13):1173-1198.

Zollinger SA, Brumm H. 2011b. The Lombard Effect. Current Biology 21(16):R614-R615.

Zuberbühler K. 2019. Evolutionary Roads to Syntax. Animal Behaviour.

Zürcher Y, Burkart JM. 2017. Evidence for Dialects in Three Captive Populations of Common Marmosets (*Callithrix Jacchus*). International Journal of Primatology 38(4):780-793.

Zürcher Y, Willems EP, Burkart JM. 2019. Are Dialects Socially Learned in Marmoset Monkeys? Evidence from Translocation Experiments. PloS one 14(10):e0222486.

Appendix -

Reprint of relevant co-authored publication

BIOLOGICAL REVIEWS
Cambridge Philosophical Society
Biol. Rev. (2017), pp. 000–000.
doi: 10.1111/brv.12382

The function and mechanism of vocal accommodation in humans and other primates

Hanna Ruch[1]*, Yvonne Zürcher[2] and Judith M. Burkart[2]

[1] *University Research Priority Program Language and Space, University of Zurich, 8032 Zürich, Switzerland*
[2] *Department of Anthropology, University of Zurich, 8057 Zürich, Switzerland*

ABSTRACT

The study of non-human animals, in particular primates, can provide essential insights into language evolution. A critical element of language is vocal production learning, i.e. learning how to produce calls. In contrast to other lineages such as songbirds, vocal production learning of completely new signals is strikingly rare in non-human primates. An increasing body of research, however, suggests that various species of non-human primates engage in vocal accommodation and adjust the structure of their calls in response to environmental noise or conspecific vocalizations. To date it is unclear what role vocal accommodation may have played in language evolution, in particular because it summarizes a variety of heterogeneous phenomena which are potentially achieved by different mechanisms. In contrast to non-human primates, accommodation research in humans has a long tradition in psychology and linguistics. Based on theoretical models from these research traditions, we provide a new framework which allows comparing instances of accommodation across species, and studying them according to their underlying mechanism and ultimate biological function. We found that at the mechanistic level, many cases of accommodation can be explained with an automatic perception–production link, but some instances arguably require higher levels of vocal control. Functionally, both human and non-human primates use social accommodation to signal social closeness or social distance to a partner or social group. Together, this indicates that not only some vocal control, but also the communicative function of vocal accommodation to signal social closeness and distance must have evolved prior to the emergence of language, rather than being the result of it. Vocal accommodation as found in other primates has thus endowed our ancestors with pre-adaptations that may have paved the way for language evolution.

Key words: vocal accommodation, phonetic accommodation, linguistic alignment, vocal plasticity, vocal learning, vocal control, Lombard effect, language evolution, non-human primates, humans.

CONTENTS

* Address for correspondence (E-mail: hanna.ruch@uzh.ch).

I. INTRODUCTION

Vocal control and vocal learning are fundamental prerequisites for the evolution of spoken language (e.g. Fitch, 2000; Tomasello, 2008; Lemasson, Ouattara & Zuberbühler, 2013; Zuberbühler, 2015). When acquiring speech, children learn from their caretakers and their social environment how to produce and combine sounds into meaningful utterances (Kuhl, 2004). To understand the phylogenetic origin of human traits and behaviour, non-human primates (henceforth primates) are crucial because they are our closest relatives. In the case of language evolution, however, they may be less informative (Fischer, 2017) given their apparent lack of vocal production learning (Janik & Slater, 2000; see Fig. 1 for different types of vocal learning). As in most mammals, primates typically do not rely on learning to acquire their vocal repertoires. Call repertoires are highly similar across populations of the same species (Egnor & Hauser, 2004), and immatures do not match their vocal output to a template, e.g. the mother's vocalization, as human children do when acquiring new words. The idea that primate vocalizations may not be crucial to understanding language evolution was further supported by reports suggesting that primates have more voluntary control over their gestures than over vocalizations (e.g. Tomasello, 2008; Liebal *et al.*, 2013; but see Schel *et al.*, 2013).

However, several studies have shown that some degree of vocal production learning nevertheless occurs at least in some primates. Learning of new call types (lexical learning, Fig. 1) has not been documented to date, but increasing evidence suggests that some primate species are able to modify existing call types (vocal accommodation, Fig. 1). For instance, several species have been shown to modify their calls in a noisy environment (e.g. de la Torre & Snowdon, 2002; Brumm & Zollinger, 2011; Hotchkin & Parks, 2013) or to change existing call types after changes in their social environment (e.g. Snowdon & Elowson, 1999; Lemasson & Hausberger, 2004). We refer to the former as environmental accommodation, and to the latter as social accommodation (Fig. 1).

There is some disagreement on how to classify and interpret findings of vocal accommodation (see Fischer, Wheeler & Higham, 2015; Watson *et al.*, 2015*b*). Here (see Fig. 1), we follow Fischer *et al.* (2015) and Fischer (2017) who stress the necessity to distinguish between different forms of vocal production learning because there is no *a priori* reason to expect that they are the result of the same mechanistic processes or have the same function. Rather, we propose that studying vocal accommodation as a separate phenomenon has the potential to shed light on several open questions concerning language evolution. First, identifying the mechanisms underlying vocal accommodation (and how they differ from lexical learning) may provide fruitful insights regarding voluntary control over vocalizations in primates, which is a precondition for the emergence of language (Tomasello, 2008; Lemasson *et al.*, 2013). Second, understanding the ultimate function of vocal accommodation and comparing it to humans may reveal which communicative functions of vocalizations are already present in non-linguistic species, and thus help to identify functions that are an exclusive consequence of language. Third, studying the distribution and variation of accommodation across different species is a promising route to understanding the phylogenetic origin of the different elements of language.

This review aims to address these issues by bringing together separate research traditions concerned with the subtle modification of vocal output (i.e. vocal accommodation): linguistics and psychology on the one hand, and primatology and animal behaviour on the other. In contrast to primates, the investigation of accommodation in humans has a long-standing tradition in psychology and linguistics, and the models of social accommodation developed in the human literature may provide a useful framework to organize the increasing body of evidence on vocal accommodation in primates, while at the same time maintaining the biologically fundamental distinction between proximate and ultimate causation of behaviour. Following Tinbergen (1963), it is crucial to distinguish between the proximate underlying mechanisms of a behaviour (i.e. concerning the causal, mechanistic cognitive processes that result in accommodation, e.g. whether accommodation is caused by voluntary processes) and its ultimate adaptive function (i.e. concerning the adaptive benefits of accommodation for survival and reproduction). Importantly, the proximate and the ultimate level of causation are independent dimensions. For instance, if it can be demonstrated that proximately, an individual would voluntarily accommodate to the call of a specific group member because it feels the need to be more similar to a friend, this does not imply that this individual is representing the ultimate goal of maintaining strong bonds with particularly valuable partners because this is associated with increased health and reproductive success (see Silk, 2014, for the link between social bonds and fitness).

In the human literature, there are two predominant models concerned with accommodation. The Interactive Alignment Model (IAM; Pickering & Garrod, 2004) from the cognitive-psychology and psycholinguistics tradition mainly emphasizes what biologists refer to as the proximate level of explanation, whereas the Communication Accommodation Theory (CAT; Giles, Coupland & Coupland, 1991; Giles & Baker, 2008) from the social-psychology and sociolinguistics tradition is more closely linked to functional and ultimate aspects. Based on the combination of these models, we first provide a new theoretical framework that allows studying and organizing the increasing evidence for vocal accommodation in animals, and analysing the distribution of accommodation and underlying processes across species. We then review instances of accommodation in humans and other species and use the framework to situate them, and compare them as to their underlying mechanism and ultimate function. A particular focus will be on primates because understanding vocal accommodation under the constraints of this specific lineage

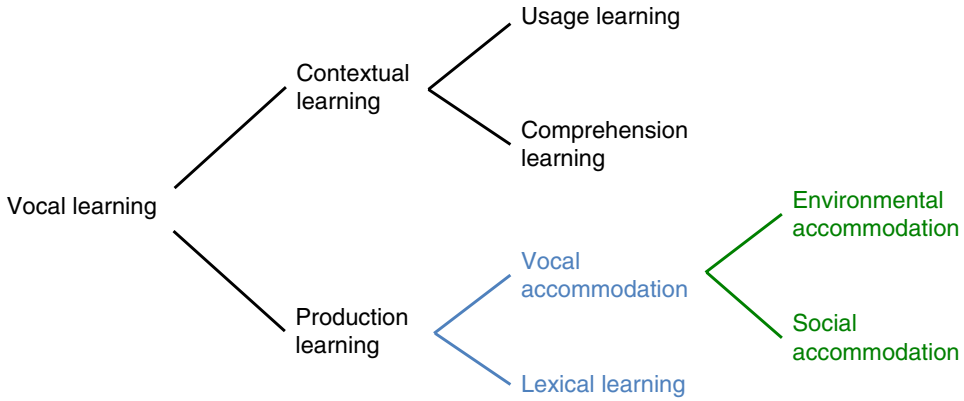

Fig. 1. An overview of different types of vocal learning based on Janik & Slater (2000) and Fischer *et al.* (2015). Contextual learning refers to a situation where 'a pre-existing signal comes to be associated with a new context' (Janik & Slater, 2000, p. 2), and thus does not involve the modification of a call. Contextual usage learning refers to a signaller learning to use a call with a new context, and contextual comprehension learning to a receiver learning to associate a pre-existing call to a new context. For instance, infant vervet monkeys have to learn to produce predator alarm calls only to real threats (usage learning), and to learn to react in an appropriate way to the alarm calls of a conspecific (comprehension learning; Seyfarth & Cheney, 1997). Production learning refers to the modification of the form of a signal or call (Janik & Slater, 2000). Further ramification within vocal production (blue) is based on Fischer *et al.* (2015) who distinguish subtle modifications of existing call types (vocal accommodation) from the acquisition of new ones (lexical learning). Vocal accommodation can be subdivided further into social accommodation and environmental accommodation (green; see text for details).

is arguably most relevant concerning the implications for language evolution. We finally discuss our findings with regard to mechanism, function and evolutionary history of distinct elements of language and highlight promising routes for future research on accommodation within linguistics and biology.

II. A THEORETICAL FRAMEWORK FOR VOCAL ACCOMMODATION

In the IAM (Pickering & Garrod, 2004), alignment (i.e. speakers becoming more similar to or converging with each other) refers to a particular case of coordination between interlocutors. It takes place at the conceptual as well as at all linguistic levels (i.e. the lexical, semantic, morphosyntactic, and phonetic levels) and can be observed as increasing similarity between interlocutors over time in their word choice, syntactic constructions, and pronunciation.

Pickering & Garrod (2004) suggest that the process of alignment is fundamentally based on a priming mechanism, i.e. a tight, automatic, and bidirectional relationship between speech perception and production. In a dialogue, one listener perceives and comprehends what is produced by his interlocutor, and *vice versa*. As a consequence, the link between perception and production within a speaker is constantly activated. It is assumed that the channels between perception and production are automatic, similar in nature to the link between perception and action (see Dijksterhuis & Bargh, 2001) that has been suggested to play a crucial role in imitation of other types of behaviours. The exact nature and tightness of the relationship between speech perception and speech production has been addressed in several theoretical approaches and is still a matter of debate (see Liberman & Mattingly, 1985; Fowler & Galantucci, 2005; Pardo, 2012; Pickering & Garrod, 2013; Beddor, 2015; Skipper, Devlin & Lametti, 2017).

The model assumes that alignment between speakers is an important, automatic mechanism in successful communication. However, human accommodation is also affected by social variables. For instance, a positive attitude towards the social group of the interaction partner favours alignment or convergence (see Section III). Such social variables are not the primary focus of the IAM and may indeed seem problematic for this automatic approach. Accordingly, in an extension of the model, Gambi & Pickering (2013) argue that social effects can be understood in terms of higher exposure to preferred social partners and their speech. Attitudes towards a social partner or group thus would affect accommodation only indirectly because they are correlated with the amount of exposure to these particular social partners and with their way of speaking, which leads to more precise forward models (simulations) and, consequently, to more convergence. Vocal convergence is thus seen to occur not due to intention or as a conversation strategy, but as a by-product of the internal and automatic mechanisms of speech perception and comprehension. Taking these extensions into account, the automatic IAM framework is able to predict a variety of, albeit not all, social effects on the degree of convergence (see below).

Communication Accommodation Theory (Giles *et al.*, 1991; Giles & Baker, 2008) addresses the relationships between language, social interaction, and social evaluation. Giles *et al.* (1991) assume that speech is a kind of social behaviour in which similarity/dissimilarity is particularly salient. According to CAT, convergence can be understood

as a reflection of a speaker's need for socially identifying or integrating with their interlocutor or the interlocutor's social group (Giles *et al.*, 1991). The function of convergent vocal accommodation (i.e. becoming more similar) is thus, through similarity in behaviour, to gain the interlocutor's liking (see similarity attraction theory; Byrne, 1971) and a way of expressing social closeness. Divergent vocal accommodation (i.e. becoming less similar), on the other hand, is a way of expressing social distance. Originally called Speech Accommodation Theory (SAT), CAT aimed at explaining language choice in intergroup communications (Giles & Powesland, 1975). Over the years, the theory was refined and applied to a variety of communication behaviours beyond language. These ideas are based on socio-psychological research that showed a relationship between the perceived behavioural similarity of a person and the ascription of positive attributes to this person, such as attractiveness or interpersonal involvement (Giles *et al.*, 1991). In CAT, accommodation is often referred to as a communicative strategy (e.g. in Giles & Baker, 2008), implying that conscious and intentional processes are involved. However, convergence or divergence need not necessarily be the result of a conscious and intentional process: 'It seems (...) that a speaker's goals may be more or less overtly represented and that speech adjustments cannot uniformly be taken as indicative of wholly intentional orientations' (Giles *et al.*, 1991, p. 25), a view that is based on inconsistencies between speakers' self-reported and actual language behaviour (Bourhis, 1984).

To summarize, CAT considers convergence or divergence as communication strategies and a way of expressing social closeness or distance, which can be more or less voluntary. In the IAM framework, convergence is seen as a result of automatic processes necessary for speech comprehension, i.e. the automatic and unintended perception–production link that enhances mutual understanding of interlocutors. An important difference, however, is that IAM can not provide an explanation for divergent forms of vocal accommodation.

Any instance of vocal accommodation can thus be classified with regard to its function, and to whether it can be understood in terms of a fully automatic mechanism such as a perception–production link, as shown in Fig. 2. Unlike IAM and CAT which exclusively focus on social accommodation, we also include vocal adjustments in response to environmental factors to reveal a more complete picture of vocal accommodation and its underlying mechanisms across species.

The framework in Fig. 2 allows the classification of instances of vocal accommodation from the literature regarding their underlying mechanism and their ultimate function. The first dimension (MECHANISM; columns in Fig. 2) refers to whether or not fully automatic processes are likely to account for findings of vocal accommodation (IAM), the second dimension (FUNCTION; rows) to whether accommodation occurs to ensure signal transmission (Lane & Tranel, 1971; environmental accommodation) or to express social closeness or distance (social accommodation; CAT).

Category A includes instances of environmental accommodation that can be traced back to automatic mechanisms (e.g. the Lombard effect, an increase in voice amplitude when exposed to a loud environment; Brumm & Zollinger, 2011). However, as evident in category B, some cases of environmental accommodation cannot be readily explained by an automatic mechanism, e.g. when individuals increase the amplitude of their call when exposed to low-amplitude calls which indicate a partner is further away (Choi, Takahashi & Ghazanfar, 2015).

Category C includes situations of vocal accommodation which can be explained by an underlying automatic mechanism such as a link between vocal perception and production, but in which the extent of convergence is influenced by social factors (e.g. more convergence between partners with a closer social relationship). Category D refers to adjustments in vocal behaviour that cannot be explained through such an automatic mechanism, and are at the same time contingent on social factors (e.g. divergence between members of different social groups).

Whereas categories C and D allow inferences regarding the social function of accommodation, categories B and D are particularly intriguing with regard to the question of whether a species is able to exert voluntary control over vocalizations, which has been argued to be responsible for the apparent lack of lexical learning in primates (Egnor & Hauser, 2004; Hammerschmidt & Fischer, 2008; Fischer, 2017; but see Hage & Nieder, 2013; Snowdon, 2017).

We will now use this framework to review findings of vocal accommodation in humans, non-human primates, and other animals and discuss them concerning the two dimensions, that is, their underlying mechanism and their ultimate function.

III. VOCAL ACCOMMODATION IN HUMANS

In addition to phonetic accommodation (subtle shifts in pronunciation), accommodation in humans can also be observed as categorical switches from one language to another in bilingual speakers (Giles, Taylor & Bourhis, 1973), shifts between a regional and a standard accent (Giles, 1973), in syntax (Branigan *et al.*, 2007; Healey, Purver & Howes, 2014), and in lexical choice (e.g. Brennan & Clark, 1996). Here, to enable comparison with the primate literature, we exclusively focus on phonetic accommodation. With more-sophisticated methods of analysis and signal processing, the focus of recent investigations lies primarily on the shifts in acoustic parameters, which is paralleled in studies with primates and thus offers the opportunity to integrate the findings from these separate research traditions.

Research on vocal accommodation in humans distinguishes between short-term accommodation (i.e. over minutes to hours) and long-term accommodation (over months or years), and it is commonly assumed that long-term accommodation is based on repeated short-term accommodation (Trudgill, 1986; Auer & Hinskens, 2005; Nguyen

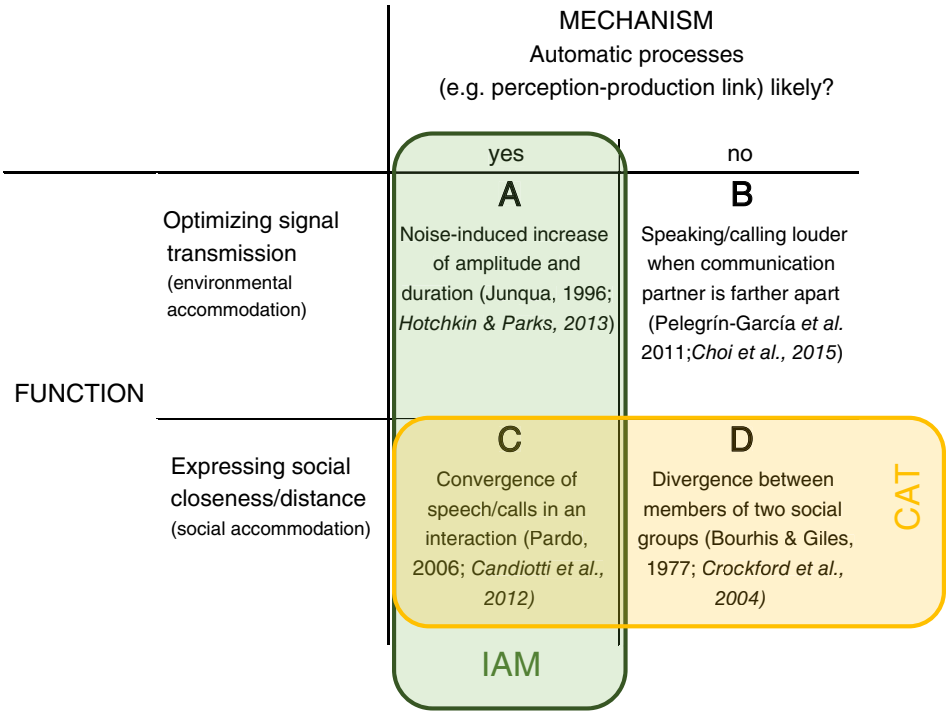

Fig. 2. Framework for studying vocal accommodation, informed by the psycholinguist/cognitive psychologist Interactive Alignment Model (IAM) and socio-psychological Communication Accommodation Theory (CAT). IAM makes a strong assumption at the proximate level, i.e. that accommodation is the result of an automatic perception–production link. CAT makes a strong assumption at the ultimate level, i.e. that accommodation serves a strategic communication function. References in *italics* refer to examples from non-human primates (see text for further details).

& Delvaux, 2015; see also Section VI). Methodologically, phonetic accommodation has been investigated with dialogue studies and shadowing tasks. Dialogue studies analyse recorded dialogues between speakers whereas in so-called shadowing tasks a speaker's baseline productions are compared against their speech after having listened to a model talker's speech over headphones (Goldinger, 1998; Shockley, Sabadini & Fowler, 2004; Babel, 2012). Herein we only include studies in which the participants were *not* explicitly asked to imitate and we therefore do not take into account so-called 'imitation studies' (see Dufour & Nguyen, 2013, for a comparison between an imitation and a shadowing task). We further include a third type of studies, namely perturbation experiments on speech in noise. In these studies, the auditory feedback of the speakers is disturbed by playing noise or babbling back to them (for reviews, see Junqua, 1996; Hotchkin & Parks, 2013; Cooke *et al.*, 2014).

Perturbation experiments thus correspond to environmental accommodation and most instances can readily be explained by an automatic mechanism (category A in Fig. 2). In a long research tradition on the so-called Lombard effect, it has been shown that when speaking in noise, participants increase pitch, amplitude, and duration of their utterances (Junqua, 1996). In perception experiments, speech recorded in noise (i.e. Lombard speech) was more intelligible than speech recorded in silence (i.e. normal speech) when mixed with noise at an equal signal-to-noise ratio (Dreher & O'Neill, 1957; Summers *et al.*, 1988; Lu & Cooke, 2008), thus supporting that environmental accommodation optimizes signal transmission. More recently, humans have been shown to be sensitive to the frequency bands of the noise they are exposed to. In noise containing speech-specific frequencies, participants increased amplitude, fundamental frequency (F0; frequency of glottal pulses, perceived as pitch) and duration; this did not take place when the speech-specific frequencies of the noise were filtered out (Stowe & Golob, 2013). Likewise, Garnier & Henrich (2014) found that speakers shifted several acoustic parameters to more quiet regions of the spectrum. These findings support the view that speakers' adjustments in adverse acoustic conditions serve to maintain signal transmission, and, ultimately, communication.

Some phonetic changes in speech output due to environmental factors cannot readily be explained by an automatic mechanism, such as when speech production patterns diverge from what is perceived (B in Fig. 2). For instance, subjects increased vocal intensity and F0 when describing a route on a map to an experimenter who was further away from them (Pelegrín-García *et al.*, 2011). If speakers simply matched their own production to the quieter sound they perceive, they should decrease rather than increase vocal intensity. Shih & Lu (2015) found that in the tone language Mandarin, duration, amplitude, F0 maximum, and onset F0 increased with increasing distance between talker and listener. However, the shape of the

F0 contour, which is relevant for lexical tone perception in Mandarin, was held constant across talker-to-listener distances, thus ensuring speech comprehension by respecting the language-specific needs of the listener.

Convergence towards a model talker in shadowing tasks can readily be explained with an automatic mechanism in the sense that the perceived speech affects speech production, leading to increased similarity between the perceived and the produced speech. In a seminal study, Goldinger (1998) found spontaneous (uninstructed) imitation in isolated words. Subsequent research using shadowing tasks (e.g. Shockley *et al.*, 2004; Nielsen, 2011; Walker & Campbell-Kibler, 2015) found that subjects pronounced isolated words in a more similar way to a model talker after having been exposed to her speech. Since these effects occur in socially impoverished environments (playbacks instead of a human interaction partner), they are arguably not motivated by social factors (A in Fig. 2). An alternative view, however, is that any linguistic stimulus may represent a social stimulus that triggers convergence even in the absence of an interlocutor.

Some shadowing studies included social variables to test the extent to which accommodation is mediated by social factors. For instance, Babel (2010) found that accommodation of New Zealand subjects towards an Australian model talker was positively affected by their implicit bias toward Australia, and Yu, Abrego-Collier & Sonderegger (2013) found that the degree of convergence towards a model talker was higher for listeners with a positive attitude towards the narrator, and that the personality trait 'openness' of the participants was correlated with their degree of convergence. These results provide evidence for the effect of social factors in phonetic accommodation (category C in Fig. 2), but they can also readily be understood in terms of increased exposure. For instance, individuals may simply be more attentive towards a preferred narrator, and more-open individuals may be more likely to pay attention to the phonetic details in the speech of a model talker (Yu *et al.*, 2013).

An automatic mechanism modulated by social factors (C in Fig. 2) can also account for convergence between speakers in a dialogue, and for long-term accommodation between persons who interact frequently over a longer time period. Coupland (1984) recorded and analysed conversations between a travel agent and 51 of her clients. Depending on the standardness of the client's English, the travel agent herself used more local (i.e. less standard) forms as well. The travel agent's convergence towards her customers can thus be understood as an attempt to increase similarity between herself and her customers by using a more or less marked local accent – which in British English is associated with social class (Trudgill, 2002). Schweitzer & Lewandowski (2014) investigated the relationship between phonetic accommodation in dialogues and mutual liking/perceived competence between unacquainted speakers of German. Overall, dialogue partners converged towards each other in vowel quality (as inferred from formant measures), and liking one's interaction partner positively affected convergence. Participants who rated their dialogue partner as more competent, however, were less-strongly influenced by their vowel quality. The authors suggest that speakers might be aware that too much convergence can be evaluated negatively, and therefore converge to a lesser degree to partners they perceive as more competent. De Looze *et al.* (2014) analysed 40 telephone conversations between native speakers of Japanese. They found that similarity in pitch, amplitude, and speech rate did not increase continuously over time, but did so between turns, and therefore evolved dynamically during a conversation. Similarity in pitch and amplitude significantly enhanced ratings by independent listeners regarding conversation fluency, balance, involvement and mutual liking of the speakers. Comparable findings were obtained by Gregory, Dagan & Webster (1997) who observed that more-similar dialogue excerpts were rated more positively on a social (e.g. friendly/unfriendly) and a power-related dimension (e.g. active/passive; dominant/submissive) by independent listeners.

In a long-term study, Pardo *et al.* (2012) found that perceived similarity of words and acoustic vowel quality of college roommates increased over the academic year. Roommate pairs who reported a greater closeness with each other also showed a higher degree of convergence. Here, as in other studies (Goldinger, 1998; Pardo *et al.*, 2013*b*; Walker & Campbell-Kibler, 2015; Abel & Babel, 2016), similarity judgments by independent listeners were only partly consistent with acoustic measures. This can be attributed to the fact that listeners use multiple cues rather than just the measured acoustic parameter when asked to judge the global similarity of speech excerpts (Pardo *et al.*, 2012).

Evans & Iverson (2007) analysed the speech of 25 young adults from Northern England before and at three time points after they had moved to Southern England. The majority of the speakers were found acoustically to shift the typically northern vowels towards a more southern accent and their speech was also rated to sound increasingly more southern-accented over time. Sancier & Fowler (1997) analysed voice onset time (VOT) of voiceless stops (i.e. /p, t, k/) in a bilingual speaker of Portuguese and English. These two languages differ in VOT, with English having a long VOT (i.e. aspirated stops) in most phonetic contexts, and Portuguese having a short VOT (i.e. unaspirated stops). After the speaker had spent a couple of months in Brazil, her VOT was shorter in both languages, and it was longer after a comparable stay in the USA.

Taken together, these studies suggest that phonetic convergence is a widespread phenomenon in humans which can readily be explained by an automatic perception–production link in combination with variation in exposure accounting for social effects. Individual differences in the extent of accommodation, which are frequently reported for humans (e.g. Evans & Iverson, 2007; MacLeod, 2012; Yu *et al.*, 2013), can occur for several reasons and do not necessarily challenge the idea of an automatic mechanism. They may be, for instance, due to individual differences in perception (Namy, Nygaard & Sauerteig, 2002), or due to different degrees of attention towards phonetic details of

speech, both of which would lead to differences in the input for the perception–production mechanism (Yu *et al.*, 2013; Abel & Babel, 2016).

Some findings are more difficult to understand on the basis of a perception–production link and frequency of exposure (category D in Fig. 2). In some studies, for instance, some groups of subjects did not converge to each other despite massive exposure (Kim, Horton & Bradlow, 2011). In dialogue studies by Pardo (2006) and Pardo *et al.* (2013*a*), the communicative role of the speakers affected the degree of convergence in vowel quality and speech rate to their interaction partners. In particular, the information givers converged more to their interaction partners than the information receivers – although information givers were exposed to their interaction partner's speech for *less* time than *vice versa*. In Pardo *et al.* (2013*a*), this asymmetric pattern even held after participants switched roles. In contrast to the prediction of IAM, it thus seems that the partner who was more interested in the information transfer converged more, rather than the partner who had higher exposure to the interactant's speech. In this example, accommodation can be interpreted as an attempt of givers to promote affiliation with their interaction partners (Pardo *et al.*, 2013*a*).

In a longitudinal study of speakers from southern Switzerland who moved to Berne (Werlen & Schlegel, 2006), five out of 18 speakers used *less* Bernese pronunciation variants two years after their relocation than shortly afterwards and therefore had to some extent reversed their convergence towards Bernese German. Thus, unlike as predicted by an automatic perception–production link, increased exposure did not automatically lead to increased convergence. At the same time and unlike in the study on British English discussed above (Evans & Iverson, 2007), only two out of 18 participants showed a clear increase in the use of Bernese pronunciation variants over time. Participants with a very low degree of convergence also reported very positive attitudes towards their home region, less-positive feelings towards Berne, and plans to return. In contrast to the UK, dialects are considered high prestige in German-speaking Switzerland and are spoken in all everyday situations (Christen, Glaser & Friedli, 2010). Consequently, there might be less communicative need and social pressure to give up one's own regional dialect (see Glaser & Bart, 2015). Another example that is hard to explain by an automatic perception–production link and frequency of exposure comes from long-term accommodation in the British journalist and radio broadcaster Alistair Cooke. After emigrating to the USA he first converged in terms of vowel quality towards the American English accent, but in later life he shifted back to his British English accent despite massive accumulated exposure to American English (Reubold & Harrington, 2015). Although age-related factors (i.e. decreasing cognitive and linguistic control) could not be totally excluded, the authors observed that Cooke's accent reversion fell into a time period in which he also showed a less-positive attitude towards American society. The British broadcaster's accent shift was thus likely an expression of his less favourable attitude towards the USA. These examples suggest that, beyond the mere effect of exposure to a linguistic variety, other factors such as attitude towards a dialect or the receiving community's experience with regional variation play a role in long-term phonetic accommodation. However, more research involving larger sample sizes is needed to confirm the role of attitudes in long-term accommodation, and to understand what other factors may inhibit or even reverse long-term accommodation.

Arguably most difficult to explain with an automatic perception–production link, finally, are cases in which speakers diverged, i.e. became linguistically *less* similar to their dialogue partners. Bourhis & Giles (1977) exposed two groups of English-speaking adults from Wales to a tape-recorded statement of a standard British English speaker who questioned the status and the future of the minority language Welsh. One group studied Welsh privately ('integrative learners'), the other group for business reasons ('instrumental learners'). The integrative learners were found to respond to the statement in a more marked Welsh-accented English, thus diverging from the British English speaker, while the instrumental learners attenuated their Welsh accent, converging towards standard British English. The divergent behaviour can be understood as an attempt to increase intergroup differences by making their Welsh identity salient. Schweitzer & Lewandowski (2013) observed overall divergence effects for articulation rate in spontaneous dialogues between unacquainted speakers. However, the direction of accommodation was correlated with mutual liking. Speakers who reported liking each other (as assessed in a post-dialogue questionnaire) were more likely to converge with each other in terms of articulation rate.

Together, these findings indicate that in humans, the general direction of accommodation – divergence or convergence – is contingent on social distance between partners and serves either to reinforce similarity and closeness in the case of convergence, or distinctiveness in the case of divergence, and is thus compatible with CAT. However, these social functions do not necessarily imply that the phenomenon cannot be captured by an automatic mechanism (see Babel, 2012). Accommodation, in particular convergence, can result from purely automatic processes and nevertheless have a distinct social function. Vocal divergence, however, cannot be explained by a simple perception–production link and thus may require additional control mechanisms.

IV. VOCAL ACCOMMODATION IN ANIMALS

(1) Vocal accommodation in primates

We now turn to primates (Table 1), where several studies have found evidence for environmental accommodation. For instance, olive baboons (*Papio anubis*) adapted their grunt vocalizations to environmental conditions producing longer calls in a closed than in an open habitat (Ey *et al.*, 2009),

Table 1. Summary of reviewed studies on vocal accommodation in animals

		MECHANISM Automatic processes?		FUNCTION		
Order	Species	Yes	No	Optimizing signal transmission	Signalling social distance/ closeness	References
Primates	Chimpanzee (*Pan troglodytes*)	X			X	Mitani & Gros-Louis (1998) and Watson *et al.* (2015*a*,*b*)
	Chimpanzee		X		X	Crockford *et al.* (2004)
	Olive baboon (*Papio anubis*)	X		X		Ey *et al.* (2009)
	Diana monkey (*Ercopithecus diana*)		X	X		Candiotti, Zuberbühler & Lemasson (2012)
	Diana monkey	X			X	Candiotti *et al.* (2012)
	Japanese macaque (*Macaca fuscata*)	X			X	Sugiura (1998) and Lemasson *et al.* (2016)
	Campell's monkey (*Cercopithecus campbelli*)	X			X	Lemasson *et al.* (2011) and Lemasson & Hausberger (2004)
	Cotton-top tamarin (*Saguinus oedipus*)	X		X		Hotchkin, Parks & Weiss (2015)
	Pygmy marmoset (*Cebuella pygmaea*)	X		X		de la Torre & Snowdon (2002)
	Pygmy marmoset	X			X	Snowdon & Elowson (1999) and de la Torre & Snowdon (2009)
	Common marmoset (*Callithrix jacchus*)	X		X		Brumm *et al.* (2004)
	Common marmoset		X	X		Choi *et al.* (2015) and see also Eliades & Wang (2012)
	Common marmoset	?			X	Y. Zürcher & J.M. Burkart (in preparation)
	Wied's black-tufted-ear marmoset (*Callithrix kuhlii*)	?			X	Rukstalis, Fite & French (2003)
Other mammals	Bats [e.g. greater horseshoe bat (*Rhinolophus ferrumequinum*)]	X		X		e.g. Hage *et al.* (2013*b*)
	Bats [e.g. greater spear-nosed bat (*Phyllostomus hastatus*), greater sac-winged bat (*Saccopteryx bilineata*)]	X			X	reviewed in Knörnschild (2014)
	Right whale (*Eubalaena glacialis*)	X		X		e.g. Parks *et al.* (2010)
	Elephant seal (*Mirounga leonina*)	X			?	Sanvito, Galimberti, & Miller (2007)
Birds	Nightingale (*Luscinia megarhynchos*)	X		X		e.g. Brumm & Todt (2002)
	Black-capped chickadee (*Poecile atricapillus*)	X			X	Nowicki (1989)
	Budgerigar (*Melopsittacus undulatus*)	X			X	Farabaugh, Linzenbold, & Dooling (1994)
	American goldfinch (*Carduelis tristis*)	X			X	Mundinger (1970, 1979)
	Pine siskin (*Spinus pinus*)	X			X	Mundinger (1970, 1979)
	European siskin (*Spinus spinus*)	X			X	Mundinger (1970, 1979)

and wild pygmy marmosets (*Cebuella pygmaea*) produced calls that fell in relatively quiet regions of the frequency spectra of their respective habitats, which reduces interfering noise (de la Torre & Snowdon, 2002). In response to increased background noise, captive common marmosets (*Callithrix jacchus*) increased their median sound level as well as call duration (Brumm *et al.*, 2004), and cotton-top tamarins (*Saguinus oedipus*) modified several frequency parameters of their calls (Hotchkin *et al.*, 2015). Lombard effects are widespread not only in primates but also in other animals (see Section IV.2) and are generally attributed to fully automatic mechanisms (reviewed in Brumm & Zollinger, 2011; Hotchkin & Parks, 2013). They thus qualify as examples of category A in Fig. 2, even though some evidence suggests that additional control processes are involved at least in some primates (see Eliades & Wang, 2012, and Section V.1).

Some findings of environmental accommodation cannot readily be explained by automatic processes alone (category B in Fig. 2), such as when the modification of the calls goes in the opposite direction of what an animal perceives. Physical features of sound, like reverberation and attenuation, allow an individual to estimate the distance to a communication partner and provide information about the transmission rate in the current environment. This information can be used to adjust calls and calling behaviour to the current condition in order to increase effectiveness of transmission. Common marmosets were found to increase the amplitude of their phee calls (a long-distance contact call) when they were responding to a conspecific who was heard from further away, ensuring that turn-taking communication would not break down (Choi *et al.*, 2015). In Diana monkey (*Ercopithecus diana*) females, the individuality of calls was more pronounced when the animals were further apart from each other and/or the visibility was low (Candiotti *et al.*, 2012). In both cases, the mechanisms of how individuals adapt their calling behaviour to the distance of a communication partner cannot be explained by an automatic perception–production link, as an individual responds to a lower call by increasing the amplitude, and therefore with the opposite of what it perceives.

Accommodation processes corresponding to C in Fig. 2, i.e. those which can be fully explained by an automatic perception–production link and have a social function, are probably the most-common case of social accommodation in primates. They are typically observed as vocal convergence between social partners. As in humans, convergence might both happen over a short time period, i.e. to a current communication partner, or over a longer time period, e.g. to new permanent social partners or groups. Examples of short-term accommodation have been found in different primate species. Japanese macaques (*Macaca fuscata*) adjusted the frequency range towards a playback stimulus (Sugiura, 1998). They also converged more to a vocal partner who was higher in the hierarchy compared to themselves, and low-ranking individuals generally showed higher variability in their calls (Lemasson *et al.*, 2016). Female Diana monkeys showed context-specific vocal accommodation: when engaging in a conversation, their calls became more similar than when two individuals were calling independently (Candiotti *et al.*, 2012). Chorusing, pant-hooting male chimpanzees (*Pan troglodytes*) also converged towards their current partners, which arguably could not be explained by similar affective states (Mitani & Gros-Louis, 1998). These studies suggest that individuals of several primate species are able to accommodate vocally to specific partners over a short time period.

Long-term accommodation has mainly been observed during the process of pair and group formation and might serve to strengthen pair bonds and group identity. Captive pygmy marmosets modified their call structure when paired with a new mate. Snowdon & Elowson (1999) showed that in three out of four newly formed pygmy marmoset pairs the individuals became more similar in their call structures after pairing than they were before; the fourth pair who did not accommodate had very similar calls already before pairing. After the pair-formation process and the initial accommodation, individual calls varied over time, but the difference between the members of a pair stayed the same, since pairs changed in the same direction. Differences in call structure were found between populations of wild pygmy marmosets (de la Torre & Snowdon, 2009), and between groups of captive common marmosets (Zürcher & Burkart, 2017). These differences in the wild could be due to convergence among group members, even though ecological and genetic effects cannot be completely excluded. Translocation experiments in captivity (Y. Zürcher & J.M. Burkart, in preparation), which showed accommodation between animals after pairing, strongly suggest that the between-group differences are the result of vocal accommodation.

Lemasson *et al.* (2011) investigated the influence of genetics and social bonds on the similarity of vocal structure in Campbell's monkey (*Cercopithecus campbelli*) contact calls. Campbell's monkeys produce contact calls that differ among individuals in the shape of the F0 contour. Individuals with a higher degree of social affiliation showed more similarities in their call structures. This pattern could not be explained by genetic relatedness, but most likely emerged because animals with a stronger social bond were more frequently exposed to each other's calls than individuals with a looser social bond. In Campbell's monkeys, changes in group composition also led to changes in the vocal pattern of several animals (Lemasson & Hausberger, 2004). Intriguingly, more-stable groups showed less-similar calls, which may indicate that instability in groups enhances the need for group cohesion and clear signals of social bonds, and therefore leads to more accommodation (Lemasson & Hausberger, 2004). Watson *et al.* (2015*b*) found that a group of captive chimpanzees from a Dutch safari park modified their food grunts after being introduced to a group of captive chimpanzees in a Scottish zoo. The shift in food grunts mainly occurred in the Dutch, but not the Scottish animals, and furthermore only after the newly arrived animals had established social bonds. Their results are somewhat controversial because it remains unclear whether auditory experience or declining arousal led to the change in acoustic

structure over time (Fischer *et al.*, 2015). In a subsequent analysis, Watson *et al.* (2015*a*) could rule out some of Fischer *et al.*'s (2015) critiques and furthermore found a positive correlation between a Dutch individual's integration with the Scottish group and their degree of convergence, but the results remain debated. In all these examples of convergence, an automatic perception–production link and variation in exposure frequency is sufficient as a mechanistic explanation, and the pattern of change is consistent with a social function, including the facilitation of individual recognition, increasing group stability or increasing the social bond in a pair (Tyack, 2008).

Instances of category D in Fig. 2 are rarer in the primate literature. They refer to situations where social accommodation cannot be readily explained by an automatic mechanism combined with variation in exposure to the calls of conspecifics. Crockford *et al.* (2004) found that directly neighbouring groups of wild chimpanzees had more distinct pant-hoots, whereas more distant groups (strangers) differed only randomly. Since genetic and environmental factors could arguably be excluded as alternative explanations, these findings suggest that the direct neighbours diverged from each other. This example is consistent with the function of divergence stressed by CAT, i.e. the demarcation of group distinctiveness, which is more relevant for neighbours than for complete strangers.

Rukstalis *et al.* (2003) showed that Wied's black-tufted-ear marmosets (*Callithrix kuhlii*) modified their call structure after unfamiliar animals were added to the colony room (in a separate cage without visual, but with auditory and olfactory contact). Due to this social stimulus, the animals changed different parameters of their calls. A control group of marmosets in a separate room without new individuals did not show any vocal changes. As the direction of change in the first group was not investigated in this study, further research is required to identify whether this species converges towards or diverges from newly introduced neighbours.

In sum, this overview of vocal accommodation suggests strong parallels in the function and mechanisms of vocal accommodation in humans and primates.

(2) Vocal accommodation in other animals

Vocal accommodation also occurs in other animals (Table 1). Environmental accommodation of category A in Fig. 2 has been reported for several mammals and birds (for reviews, see Brumm & Slabbekoorn, 2005; Brumm & Zollinger, 2011; Hotchkin & Parks, 2013), and more recently, in a fish species (Holt & Johnston, 2014). The Lombard effect has been documented, for instance, in greater horseshoe bats (*Rhinolophus ferrumequinum*) (e.g. Hage *et al.*, 2013*b*) North Atlantic right whales (*Eubalaena glacialis*) (e.g. Parks *et al.*, 2010), and nightingales (*Luscinia megarhynchos*) (e.g. Brumm & Todt, 2002). By contrast, category B accommodation (Fig. 2) has rarely been reported in non-primates.

Social accommodation has been observed in some non-primate species. In a longitudinal study on elephant seals (*Mirounga leonina*), Sanvito *et al.* (2007) showed that the variations of agonistic calls of dominant males became more prevalent in young males over the years in the common breeding ground, and that call variations from dominant males were more widely distributed in later years, indicating a certain involvement of vocal accommodation. Whether vocal accommodation in this example has the same social function of indicating social closeness or distance as in primates is not clear, since the juvenile males most probably did not express social closeness with the dominant males. Likely alternatives are that this case of vocal convergence can be understood in the context of sexual competition, that it is a simple by-product of frequent exposure to the dominants' calls (Sanvito *et al.*, 2007), or that it represents maturation rather than convergence. Evidence for bats showing convergence of innate calls towards conspecifics is reviewed in Knörnschild (2014). In bats, this phenomenon has been observed in both juveniles and adults, and it ultimately serves to signal group membership (Knörnschild, 2014).

Social accommodation has also been reported for some birds. For instance, Nowicki (1989) documented how the song of five captive black-capped chickadees (*Poecile atricapillus*) converged in several acoustic parameters over 3 weeks, with changes taking place even after the first week. Farabaugh *et al.* (1994) found that social interaction had an effect on whether budgerigars (*Melopsittacus undulatus*) became vocally more similar. Animals that could see each other converged over time, by recombining and imitating songs (song-type matching, see below), but also by subtly modifying acoustic properties as soon as they shared call types. Birds that were kept in visually separated cages were not found to converge. Several species of the Carduelinae family also show evidence of social accommodation. American goldfinches (*Carduelis tristis*) and pine siskins (*Spinus pinus*) accommodated certain calls to new breeding partners even across closely related species (when interbreeding in captivity), and European siskin (*Spinus spinus*) males as well as pine siskins showed vocal accommodation when introduced to other males (Mundinger, 1970, 1979). When encountering other European siskin males, two submissive males modified their calls along with more benevolent behaviours, whereas two dominant males, which where hostile towards each other, did not (Mundinger, 1970, 1979). Several Carduelinae species have furthermore been documented to show pair-specific call patterns, which could be the result of accommodation during pair formation (Mundinger, 1979).

In general, it seems that environmental accommodation is more widespread than social accommodation in non-primate species (see also Tyack, 2016). Although this distribution of findings could be the result of more research effort in one area, another likely explanation is that the mechanism permitting environmental accommodation of category A in Fig. 2 is phylogenetically older (see Section V.3).

We will briefly discuss two phenomena related to vocal accommodation: vocal matching and vocal mimicry. These are likely based on a different mechanism, but nonetheless have functions similar to social accommodation. Vocal or song-type matching takes place when an individual uses

the acoustically most-similar call of its repertoire after perceiving the call of a conspecific (King & McGregor, 2016), and therefore falls into the category of usage learning (Fig. 1). Apart from birds, for which song-type matching has been observed in several species (see King & McGregor, 2016; Sewall, Young & Wright, 2016), vocal matching has also been documented for cetaceans. For instance, Smolker & Pepper (1999) observed that male bottlenose dolphins (*Tursiops* sp.) after forming an alliance started to predominantly use the same four whistle types, and found that wild bottlenose dolphins matched their signature whistles – a call type that permits recognition of individuals and groups – to other individuals present within a distance of up to 580 m. On a much larger geographical scale, Garland *et al.* (2011, 2013) documented how new song patterns in male humpback whales (*Megaptera novaeangliae*) diffused among different populations from East Australia eastwards up to French Polynesia. Functionally, vocal matching seems more diverse than vocal accommodation. Songbirds seem to use song-type matching primarily as an aggressive behaviour, e.g. in territorial defence, and the songs of humpback whales evolved as sexual displays in the context of sexual selection (Garland *et al.*, 2011) whereas other cetaceans and parrots appear to use call-type matching to announce or increase social bonds, similar to social accommodation in primates (King & McGregor, 2016).

Vocal mimicry refers to the process by which an individual learns how to produce a non-species-specific sound, for instance, human sounds, and classifies as lexical learning (see Fig. 1). Vocal mimicry is widespread in some songbirds and parrots (reviewed in Tyack, 2016) but has also been reported in some individuals of mammals who do not usually engage in lexical learning, such as elephants [*Loxodonta africana* (Poole *et al.*, 2005); *Elephas maximus* (Stoeger *et al.*, 2012)], white whales (*Delphinapterus leucas*) (e.g. Ridgway *et al.*, 2012), harbor seals (*Phoca vitulina*) (Ralls, Fiorelli & Gish, 1985), and orangutans (*Pongo* sp.; Wich *et al.*, 2008; Lameira *et al.*, 2016). Whether, and if so how, vocal mimicry is linked to vocal accommodation is not known. Intriguingly, mimicry of anthropogenic noise is often expressed around sexual maturity despite massive exposure earlier in life (e.g. in the harbor seal and the elephant), similar to song learning in classical songbird models (Tyack, 2016). For primates, vocal matching has not been reported to our knowledge, and vocal mimicry based on vocal fold control is restricted to one captive orangutan engaging in a 'do-as-I-do' test paradigm (Lameira *et al.*, 2016). Both phenomena are thus unlikely to be part of the general primate background upon which language evolved in the hominin lineage and are therefore not discussed further below.

V. DISCUSSION

An increasing number of studies reports vocal accommodation in primates, but the implications for language evolution are still unclear (e.g. Fischer *et al.*, 2015; Fischer, 2017; Snowdon, 2017). In particular, it is not known (*i*) whether vocal accommodation in primates is based on vocal control, (*ii*) what functions accommodation serves in non-linguistic compared to linguistic species, and (*iii*) what the implications are for the evolutionary history of language emergence. A framework inspired by the most influential models of accommodation from psychology and linguistics, which at the same time respects the Tinbergian tradition of separating proximate and ultimate levels of explanation, allowed us to organize a large body of findings on accommodation in humans, primates, and non-primate animals. Based on this review, we now return to the three issues raised at the start: identifying the mechanisms underlying vocal accommodation, comparing the ultimate functions of accommodation across species, and drawing a possible evolutionary pathway of vocal production learning, a key element of language evolution.

(1) What proximate mechanisms underlie vocal accommodation in different species?

Most examples of vocal accommodation in humans and other animals can be explained by cognitively undemanding, automatic processes. For environmental accommodation, neurological research on the Lombard effect in domestic cats (*Felis catus*) (Nonaka *et al.*, 1997) and squirrel monkeys (*Saimiri sciureus*) (Hage, Jürgens & Ehret, 2006) suggests that the mechanisms for the Lombard effect are located in the brainstem. These findings are compatible with the fact that the Lombard effect is widespread in vertebrate species, and the traditional idea that primate vocalizations are mainly under subcortical control. However, at least in marmoset monkeys, not only automatic brainstem reflexes, but also cortical processes and self-monitoring are involved in the Lombard effect (Eliades & Wang, 2008, 2012). This finding is consistent with studies showing that in humans, the Lombard effect is both under automatic and voluntary control: when instructed to do so, humans can decrease, but not completely suppress the effect (Pick *et al.*, 1989), it is enhanced in content as opposed to function words (Patel & Schell, 2008), in interactive as opposed to non-interactive tasks (Amazi & Garber, 1982; Garnier, Henrich & Dubois, 2010), and it is sensitive to the linguistic function of phonetic detail (Shih & Lu, 2015).

For social accommodation, the predominant mechanistic explanation refers to an automatic link between vocal perception and production, as in particular stressed by the IAM (Pickering & Garrod, 2004). The idea that vocal perception and production at least in part share the same neural system has been supported by work on both humans (Fadiga *et al.*, 2002; Watkins, Strafella & Paus, 2003; D'Ausilio *et al.*, 2014) and songbirds (Prather *et al.*, 2008). To date, however, the exact nature of such a mechanism for accommodation in human and non-human primates remains unclear because most neurobiological work has focused (*i*) on *either* the neural bases of perception *or* production, but rarely on both, and (*ii*) on lexical learning (song learning) rather than accommodation. Although the neural mechanisms of song learning have been widely studied in songbirds and reveal

astonishing convergence with humans (e.g. Belyk *et al.*, 2016) it is not yet known to what extent vocal accommodation and lexical learning are overlapping or distinct phenomena.

Tyack (2016) summarizes additional hypotheses about potential mechanisms underlying convergence. For instance, convergence may simply result as an artifact from emotional responses to being housed with strangers and thus not involve vocal production learning at all (Owren, Amoss & Rendall, 2011). Even if in many cases emotional arousal cannot be ruled out completely, it is unlikely to account fully for the entire range of vocal accommodation phenomena observed in primates, as reviewed above. Arriaga & Jarvis (2013) suggest that a central pattern generator (CPG) located in the midbrain and/or brainstem may be responsible for call production, whereas the fine-tuning of these calls would be based on cortical input and integrated auditory pathways that modify the CPG [see Hage & Nieder, 2016 for a similar proposal].

Arriaga & Jarvis' (2013) mechanism may account for the marmoset case of the Lombard effect, and for those cases of accommodation which are more difficult to explain with fully automatic processes only, for instance when individuals *increase,* rather than *decrease* the amplitude of their calls to respond to partners that are further away (Choi *et al.*, 2015). Whether the latter can be explained fully by the regulation of arousal states *via* vocal feedback (Choi *et al.*, 2015), or whether it includes additional forms of vocal control remains to be established. Other examples difficult to explain with automatic processes include instances when individuals diverge (Crockford *et al.*, 2004) rather than converge to the vocalizations of conspecifics. Such cases are particularly intriguing from the mechanistic perspective because they may be indicative of voluntary control over vocalizations. Voluntary control over vocalizations has been argued to be one of the major differences in vocal communication between humans and primates, and is generally accepted as being necessary for language to evolve in the first place (Tomasello, 2008; Lemasson *et al.*, 2013).

The idea that voluntary control over vocalizations in primates is more widespread than traditionally thought also receives some empirical support. For instance, some primates have been reported to inhibit spontaneous answering, for instance, in anticipation of environmental noise [common marmosets (Roy *et al.*, 2011), cotton-top tamarins (Egnor, Wickelgren & Hauser, 2007)], in risky environments [e.g. pant-hoot calls in wild chimpanzees (Wilson, Hauser & Wrangham, 2007) or whisper-like behaviour in captive cotton-top tamarins (Morrison & Reiss, 2013)], or in the presence of high-ranking individuals [e.g. chimpanzee females (Laporte & Zuberbühler, 2010)]. For these latter cases, however, purely arousal-based alternative explanations still need to be ruled out. Further evidence for a certain amount of vocal control in primates comes from less-naturalistic contexts, i.e. from the orangutan imitating the vocal output of a human caregiver mentioned above (Wich *et al.*, 2008; Lameira *et al.*, 2016), from chimpanzee individuals using a species-atypical call to attract a human's attention (Hopkins, Taglialatela & Leavens, 2007), and from rhesus macaques (*Macaca mulatta*) trained to emit different call types on command (Hage, Gavrilov & Nieder, 2013*a*).

At the neurobiological level, studies with rhesus macaques using simultaneous single-cell recordings suggest a cardinal role of the monkey homologue of Broca's area in vocal planning (Hage & Nieder, 2013). Prefrontal neurons have been shown to be involved in marmoset monkeys' vocal production (Eliades & Wang, 2012; Miller *et al.*, 2015), and Taglialatela *et al.* (2008) provide evidence for activation in the chimpanzee left inferior frontal gyrus, a homologue to Broca's area in humans, during communicative gestural and vocal signalling. This activation was stronger in two chimpanzees who produced calls to catch the experimenter's attention than in two chimpanzees who only used manual gestures (Taglialatela *et al.*, 2011). Thus monkeys and apes, although unable tightly to synchronize laryngeal and articulatory movements as required for speech (see Ackermann, Hage & Ziegler, 2014; Kumar, Croxson & Simonyan, 2016; Loh *et al.*, 2016), appear to possess more vocal control than traditionally thought.

Neurobiological studies on the brain regions involved during accommodation, in particular of instances in categories B and D of Fig. 2, are still lacking and critically needed to advance our understanding of the amount of vocal control involved in accommodation. An alternative, more feasible and non-invasive way to investigate the amount of cognitive control during accommodation could be a cognitive-load approach, which predicts trade-offs between explicit mental activities that rely on the limited resource of working memory (Paas, Renkl & Sweller, 2003). Thus, if vocal accommodation is under voluntary control and therefore cognitively costly, trade-offs between accommodation and other cognitive activities that increase cognitive load are predicted. Accordingly, Abel & Babel (2016) found that in humans, convergence (as assessed by independent listeners) occurred when listeners were solving an easy, but not when they were solving a difficult collaborative task.

(2) What is the function of vocal accommodation in humans and other primates?

Most instances of vocal accommodation in primates can be readily understood based on a simple perception–production link that is part of the automatic mechanisms that have evolved for speech or vocal comprehension. Therefore, the null model for the function of vocal accommodation must be that it is a simple by-product of these mechanisms. The available data, however, suggest that vocal accommodation in humans but also in non-human animals is more than that. First, it seems unlikely that all instances of vocal accommodation can be explained fully by these mechanisms (see Section V.1), and second, the patterns of vocal accommodation observed in humans and in non-human animals, in particular primates, are consistent with an adaptive function, both in cases that can and cannot be understood in terms of a simple perception–production link.

In instances of environmental accommodation, the reviewed changes in vocalization structure are consistent with the function of improved signal transmission. Environmental accommodation can increase signal transmission in several ways, for instance *via* increased vocal amplitude or call duration in noisy environments, or if the signal is shifted to a more quiet frequency range relative to the ambient noise (Tyack, 2008). As communication *per se* is a social act and environmental accommodation enhances the chance of successful signal transmission, the ultimate function of environmental accommodation can also be considered social, although more indirectly.

In instances of social accommodation, patterns of convergence and divergence in humans and primates are in line with the predictions of CAT. In particular, convergence occurs in cases where it is important to reduce social distance and is likely to increase the chance of maintaining contact, to speed up the recognition of group members and calls, to signal social closeness and support for a group member, or to strengthen bonds between individuals and increase group stability (see Tyack, 2008; Sewall *et al.*, 2016). Convergence might also facilitate social integration and therefore allow animals to change between groups under favourable conditions (Sewall *et al.*, 2016). Divergence, by contrast, was observed in between-group contexts and to demarcate differences between individuals. Diverging from an out-group might also function to increase in-group coherence, particularly when it co-occurs with convergence within the group, as observed in chimpanzees (Crockford *et al.*, 2004). Social accommodation thus provides a likely explanation for group differences in primate vocalizations (e.g. Fischer, Hammerschmidt & Todt, 1998; Zürcher & Burkart, 2017). Likewise, in humans, accommodation may be a driving force for the emergence of new language varieties such as regional dialects, social dialects, or mixed varieties (e.g. Trudgill, 2008).

Taken together, there are surprisingly strong parallels in the function of vocal accommodation in humans and primates, which are consistent with the optimization of signal transmission and CAT and suggest an early phylogenetic origin of these functions. These results strongly suggest that this social function of accommodation already existed prior to the evolution of language.

(3) Evolutionary history and development

The most widespread form of vocal production learning seems to be environmental accommodation, and it may even be ubiquitous in species engaging in acoustic communication (Brumm & Zollinger, 2011) because once acoustic communication has evolved, the adaptive pressure to ensure signal transmission automatically follows. How reliable signal transmission is achieved nevertheless varies, depending on how the signal is produced.

In species who produce acoustic signals by pressing air through vocal folds, environmental accommodation, perhaps even stabilizing the acoustic structure of vocalizations *per se*, most likely requires some form of auditory–vocal feedback (Tyack, 2016). We can thus assume that auditory–vocal feedback allowing for environmental accommodation represents a deep homology (i.e. has a very old evolutionary origin), and is the ancestral condition upon which additional forms of vocal learning could build (Tyack, 2016). Other ways to optimize signal transmission include controlled planning of vocal exchanges by anticipating predictable periods of environmental noise, as demonstrated for marmoset monkeys (Roy *et al.*, 2011). This suggests that additional behavioural control mechanisms can be recruited for the evolutionarily old function of optimizing signal transmission in lineages such as primates, where such control mechanisms are available and can be used flexibly.

Social accommodation in the form of convergence is also widespread. Tyack (2016) suggests that the pre-existing connections between auditory input and vocal output established for environmental accommodation were co-opted to adjust call structures to conspecifics, in order to signal social closeness. Evidence for divergence, on the other hand, is rarer and has been reported for primates (Table 1).

We therefore propose the following working hypothesis: social convergence first occurred as a side-effect of environmental accommodation but soon took over an independent social function, i.e. to signal social closeness. In highly social species such as primates, where demarcating social distance plays an important role within and among complex societies, this may have paved the way for the emergence of divergent forms of accommodation. Important empirical predictions of this working hypothesis include (*i*) that vocal convergence should only be found in species that also engage in environmental accommodation based on auditory–vocal feedback, and (*ii*) that vocal divergence should only be found in species that also show evidence for vocal convergence. These may be primate or non-primate species but are predicted to be highly gregarious.

Social accommodation and its corresponding functions are thus clearly not unique to humans but are present in a variety of primates and several other animals. Evidence from primates is not restricted to great apes, our closest relatives, which would suggest a recent origin in the last common ancestor of humans and great apes some 13 million years ago (Glazko & Nei, 2003; Langergraber *et al.*, 2012). Rather, evidence is present for all major anthropoid clades, i.e. the apes, the Old World monkeys, and the New World monkeys. This may either suggest that social accommodation is a synapomorphy of anthropoid (haplorrhine) primates, placing its evolutionary origin at least 35 million years ago (Glazko & Nei, 2003), or perhaps even of mammals and birds in general. Alternatively, social accommodation may be more patchily distributed among primates and some other species, with results suggestive of absence of accommodation simply not published, or where species in which it is not expected are not studied at all. Such a patchy distribution typically emerges as a convergent response to some socio-ecological factor present in only some species. For instance, it has been argued that among primates, communicative systems may be particularly complex in the

cooperatively breeding marmoset and tamarin monkeys (Snowdon, 2001, 2013; Rukstalis *et al.*, 2003; Zuberbühler, 2011), perhaps because of their inclination towards higher levels of prosociality and their need routinely to coordinate a variety of activities (e.g. infant care, vigilance, and foraging) with other group members (Burkart & van Schaik, 2016*b*; J.M. Burkart, E. Guerreiro Martins, F. Miss & Y. Zürcher, in preparation). This cooperative breeding hypothesis would predict an over-representation of social accommodation in cooperatively breeding species (Borjon & Ghazanfar, 2014). In the primate data reviewed above, callitrichid monkeys (the only cooperatively breeding primates, together with humans) indeed seem over-represented (eight of the 18 reviewed primate studies are from callitrichids), and furthermore, social learning appears also to play a substantial role during vocal ontogeny, which is unusual for primates (reviewed in Snowdon, 2017). Additional comparative data are necessary to test systematically whether this represents a true pattern or merely results from higher research and publication effort in these species. However, since humans are the only other cooperatively breeding primates, this hypothesis has the potential to provide at least part of the explanation for why language evolved in humans, but not in any other great ape (Burkart, Hrdy & van Schaik, 2009; Burkart & van Schaik, 2016*a*).

VI. IMPLICATIONS AND FUTURE DIRECTIONS

Our review has revealed strong parallels between environmental and social accommodation in humans and primates. The finding that social accommodation serves comparable functions in individuals of different linguistic and non-linguistic species has important implications. First and foremost, it strongly suggests that these social functions are not the consequence of having evolved language, but that they are rather part of the socio-cognitive foundation upon which language was grafted during evolution. The inclusion of these social functions as antecedents, rather than as a consequence of language has implications for theories on language evolution such as approaches of cultural group selection (Richerson *et al.*, 2016).

Vocal accommodation can be seen as part of the broader phenomenon of homophily, i.e. the tendency of individuals to prefer others that are perceived to be similar to self. Homophily is pervasive in humans and likely to be involved in the maintenance of cultural differences despite frequent migration (Haun & Over, 2015), but recent studies with primates suggest that homophily may also be widespread in primates, including great apes [chimpanzees (Massen & Koski, 2014)], Old World monkeys [baboons (*Papio ursinus*; Carter *et al.*, 2015); Barbary macaques (*Macaca sylvanus*; Molesti & Majolo, 2015)], and New World monkeys [capuchin monkeys (*Cebus apella*; Paukner *et al.*, 2009; Morton *et al.*, 2015); common marmosets (Koski & Burkart, 2015)]. Related to homophily is the role of accent or dialect as a facilitator for cooperation. Cohen (2012) suggests that accent – as a socially acquired but hard-to-fake trait – may serve as a tag for cooperation between non-kin individuals: young children acquire a dialect or a local accent 'for free', but adults despite being exposed to a non-native dialect or accent for a long time (presupposing cohabitation with the respective social group) will hardly ever acquire it perfectly (Siegel, 2010). First studies indicate that in humans a regional accent may indeed serve as a criterion for cooperation (Cohen & Haun, 2013; Heblich, Lameli & Riener, 2015). Interactive research designs as well as a more diverse set of languages would be desirable to elucidate the relationship between accommodation and cooperation. For primates, the empirical prediction that ensues is that similarity at different levels, including acoustic similarity, not only increases mutual liking but also cooperation. Furthermore, more cooperative species (e.g. marmosets as opposed to macaques) are predicted to show a stronger inclination to engage in vocal convergence. Controlled experiments will be key to addressing these questions in animals, to confirm earlier, behavioural studies, and to allow for comparisons across species.

What still remains to be established is to what extent social accommodation and lexical learning are overlapping or distinct phenomena, as discussed above. But even the link between short- and long-term accommodation is not well understood. In linguistics it is commonly assumed that longer term changes in an individual are the result of repeated short-term accommodation (e.g. Trudgill, 1986; Nguyen & Delvaux, 2015), and that these changes are central to explaining language change (Bloomfield, 1933; Trudgill, 1986; Auer & Hinskens, 2005; Nguyen & Delvaux, 2015). The assumption of a relationship between short-term and long-term accommodation seems plausible and traces of convergence have been shown to last for several minutes (Pardo, 2006; Delvaux & Soquet, 2007; Dufour & Nguyen, 2013) or for days (Goldinger & Azuma, 2004). Nevertheless, despite extensive evidence for convergence, adult humans typically do not completely acquire a second dialect (Siegel, 2010), which suggests that, at least in humans, this relationship may be more multi-faceted. On the one hand, short-term accommodation itself never leads to a complete mirroring of an interaction partner's speech and does not affect all phonetic parameters to the same degree (see below). On the other hand, additional effects resulting from cultural group selection might mediate long-term accommodation. Yet another factor that may inhibit short-term accommodation accumulating into long-term accommodation is a speaker's conscious decision to refuse to acquire a new dialect or, respectively, to give up one's own. To date, empirical research on the relationship between short-term and long-term accommodation is still largely lacking [but see Sonderegger, Bane & Graff, 2017 for a comparison between day-by-day variability and longer-term time trends in individual speakers].

Likewise underexplored is whether all acoustic parameters are equally susceptible to social accommodation. For primates, some acoustic parameters such as amplitude are

likely to be easier to adjust than others that may require additional control mechanisms (Janik & Slater, 2000). If so, we will be able to refine the evolutionary question and analyse the phylogenetic origin of specific control mechanisms rather than of accommodation in a broad sense. Comparing different acoustic parameters is further relevant to understanding which are correlated and likely to be influenced by arousal (see Fischer *et al.*, 2015) or increased vocal effort, or whether they are modulated and controlled separately (see also Hotchkin & Parks, 2013).

For humans, production constraints are less likely, but instead linguistic as well as social factors may influence whether or not a specific phonetic variant is accommodated by a speaker. Nielsen (2011) found that speakers of American English imitated lengthened, but not shortened VOT of /p/, a result that was interpreted with the phonological consequences of shortening VOT in English: while increasing VOT of /p/ does not change the meaning of the word, decreasing VOT can lead to confusion of the phonemes /p/ and /b/. Apart from their function within the linguistic system, phonetic variants may also vary in the extent to which they are perceptually associated with a specific social group or stereotype (e.g. Campbell-Kibler, 2011; Foulkes *et al.*, 2011; Pharao *et al.*, 2014; Walker *et al.*, 2014). Socially highly salient parameters such as dialect differences that speakers are aware of seem to be less susceptible to convergence (Werlen & Schlegel, 2006; Babel, 2010; Walker & Campbell-Kibler, 2015). This could indicate that speakers –although overall converging – avoid adopting elements overtly associated with a different social identity to avoid convergence being perceived as ingratiation, 'dialect faking', or mocking (see Giles *et al.*, 1991). Perception experiments will be central to assessing the perceptibility of and reactions to different accommodated parameters.

VII. CONCLUSIONS

(1) With the critical input from linguistics and psychology, it has been possible to organize the continuously increasing number of vocal accommodation studies in primates and other animals, building on the knowledge of accommodation research, a field that is well established in humans, but still relatively new for primates.

(2) Our review has shown that most, but not all human and non-human primate instances of vocal accommodation can be explained with an automatic mechanism such as a link between perception and production (Pickering & Garrod, 2004). However, examples were found for both humans and non-human primates in which accommodation could not be explained by automatic processes alone (e.g. divergence from a conspecific or increasing amplitude when calling/speaking to a partner who is further away). These instances are particularly promising for studying the mechanisms underlying accommodation, and for studying vocal flexibility in animals.

(3) Strong parallels with respect to the function of vocal accommodation in humans, primates and other animals are apparent, and we propose a working hypothesis for its evolution. The functional parallels across species suggest that the communicative function of vocal accommodation to signal social closeness or social distance evolved prior to the emergence of language, rather than being the result of it.

(4) Vocal accommodation can be understood as part of the broader phenomenon of homophily, and is compatible with the idea that accent may serve as a tag for cooperation (Cohen, 2012). While children acquire a regional or social dialect for free, accommodation in adults commonly involves very subtle shifts and, even in the long term, hardly ever leads to the complete acquisition of a new dialect.

(5) Promising directions for future research include empirically testing the link between vocal accommodation (and dialects) and cooperation, both within and among species. The latter requires the development of methodological approaches to compare the extent of accommodation across species. Future directions include studying the relationship between short-term and long-term vocal accommodation as well as the range of parameters affected by convergence and divergence in different species.

VIII. ACKNOWLEDGEMENTS

The idea of a comparative approach to vocal accommodation was seeded in a joint workshop by biologists, linguists, and psychologists held in May 2015 at University of Zurich, organized by H.R., Y.Z., and J.M.B. We would like to thank Balthasar Bickel, Elodie Briefer, Volker Dellwo, Susanne Grassmann, Marta Manser, Simon Townsend, and Carel van Schaik for inspiring presentations and discussions at our workshop. We are grateful to Elvira Glaser, Ulrich Reubold, Jessica Siddins, Carel van Schaik, and two anonymous reviewers for very helpful comments on earlier versions of this paper. This work was supported by the University of Zurich's Research Priority Program Language and Space and an SNSF grant 105312-114107 to J.M.B.

IX. REFERENCES

Abel, J. & Babel, M. (2016). Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech* **60**, 479–502.

Ackermann, H., Hage, S. R. & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *Behavioral and Brain Sciences* **37**, 529–546.

Amazi, D. K. & Garber, S. R. (1982). The Lombard sign as a function of age and task. *Journal of Speech and Hearing Research* **25**, 581–585.

Arriaga, G. & Jarvis, E. D. (2013). Mouse vocal communication system: are ultrasounds learned or innate? *Brain and Language* **124**, 96–116.

Auer, P. & Hinskens, F. (2005). The role of interpersonal accommodation in a theory of language change. In *Dialect Change. Convergence and Divergence in European Languages* (eds P. Auer, F. Hinskens and P. Kerswill), pp. 1–50. Cambridge University Press, Cambridge.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society* **39**, 437–456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* **40**, 177–189.

Beddor, P.S. (2015). The relation between language users' perception and production repertoires. In *Proceeedings of the 18th International Congress of Phonetic Sciences*, Article no. 1041. Glasgow. 10–14 August 2015.
Belyk, M., Pfordresher, P. Q., Liotti, M. & Brown, S. (2016). The neural basis of vocal pitch imitation in humans. *Journal of Cognitive Neuroscience* **28**, 621–635.
Bloomfield, L. (1933). *Language*. H. Holt and Company, New York.
Borjon, J. I. & Ghazanfar, A. A. (2014). Convergent evolution of vocal cooperation without convergent evolution of brain size. *Brain, Behavior and Evolution* **84**, 93–102.
Bourhis, R. (1984). Cross-cultural communication in Montreal: two field studies since bill 101. *International Journal of the Sociology of Language* **46**, 33–47.
Bourhis, R. & Giles, H. (1977). The language of intergroup distinctiveness. In *Language, Ethnicity, and Intergroup Relations*, (ed. H. Giles) pp. 119–135. Academic Press, London and New York.
Branigan, H. P., Pickering, M. J., McLean, J. F. & Cleland, A. A. (2007). Syntactic alignment and participant role in dialogue. *Cognition* **104**, 163–197.
Brennan, S. E. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **22**, 1482–1493.
Brumm, H. & Slabbekoorn, H. (2005). Acoustic communication in noise. *Advances in the Study of Behavior* **35**, 151–209.
Brumm, H. & Todt, D. (2002). Noise-dependent song amplitude regulation in a territorial songbird. *Animal Behaviour* **63**, 891–897.
Brumm, H., Voss, K., Köllmer, I. & Todt, D. (2004). Acoustic communication in noise: regulation of call characteristics in a new world monkey. *Journal of Experimental Biology* **207**, 443–448.
Brumm, H. & Zollinger, S. A. (2011). The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour* **148**, 1173–1198.
Burkart, J. M., Hrdy, S. & van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology: Issues, News, and Reviews* **18**, 175–186.
Burkart, J. M. & van Schaik, C. P. (2016*a*). Revisiting the consequences of cooperative breeding. *Journal of Zoology*, **299**, 77–83.
Burkart, J. M. & van Schaik, C. P. (2016*b*). The cooperative breeding perspective helps pinning down when uniquely human evolutionary processes are necessary - and when not. A commentary on Richerson et al. *Behavioral and Brain Sciences* **39**, e34.
Byrne, D. E. (1971). *The Attraction Paradigm*. Academic Press, New York.
Campbell-Kibler, K. (2011). The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change* **22**, 423–441.
Candiotti, A., Zuberbühler, K. & Lemasson, A. (2012). Convergence and divergence in Diana monkey vocalizations. *Biology Letters* **8**, 382–385.
Carter, A. J., Lee, A. E. G., Marshall, H. H., Ticó, M. T. & Cowlishaw, G. (2015). Phenotypic assortment in wild primate networks: implications for the dissemination of information. *Royal Society Open Science* **2**, 140444.
Choi, J. Y., Takahashi, D. Y. & Ghazanfar, A. A. (2015). Cooperative vocal control in marmoset monkeys via vocal feedback. *Journal of Neurophysiology* **114**, 274–283.
Christen, H., Glaser, E. & Friedli, M. (2010). *Kleiner Sprachatlas der deutschen Schweiz*. Huber Frauenfeld, Frauenfeld.
Cohen, E. (2012). The evolution of tag-based cooperation in humans: the case for accent. *Current Anthropology* **53**, 588–616.
Cohen, E. & Haun, D. (2013). The development of tag-based cooperation via a socially acquired trait. *Evolution and Human Behavior* **34**, 230–235.
Cooke, M., King, S., Garnier, M. & Aubanel, V. (2014). The listening talker: a review of human and algorithmic context-induced modifications of speech. *Computer Speech & Language* **28**, 543–571.
Coupland, N. (1984). Accommodation at work: some phonological data and their implications. *International Journal of the Sociology of Language* **1984**, 49–70.
Crockford, C., Herbinger, I., Vigilant, L. & Boesch, C. (2004). Wild chimpanzees produce group-specific calls: a case for vocal learning? *Ethology* **110**, 221–243.
D'Ausilio, A., Maffongelli, L., Bartoli, E., Campanella, M., Ferrari, E., Berry, J. & Fadiga, L. (2014). Listening to speech recruits specific tongue motor synergies as revealed by transcranial magnetic stimulation and tissue-Doppler ultrasound imaging. *Philosophical Transactions of the Royal Society B* **369**, 20130418.
de la Torre, S. & Snowdon, C. T. (2002). Environmental correlates of vocal communication of wild pygmy marmosets, *Cebuella pygmaea*. *Animal Behaviour* **63**, 847–856.
de la Torre, S. & Snowdon, C. T. (2009). Dialects in pygmy marmosets? Population variation in call structure. *American Journal of Primatology* **71**, 333–342.
De Looze, C., Scherer, S., Vaughan, B. & Campbell, N. (2014). Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication* **58**, 11–34.
Delvaux, V. & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* **64**, 145–173.
Dijksterhuis, A. & Bargh, J. A. (2001). The perception-behavior expressway: automatic effects of social perception on social behavior. In *Advances in Experimental Social Psychology* (ed. M. Zanna), pp. 1–40. Academic Press, San Diego.
Dreher, J. J. & O'Neill, J. (1957). Effects of ambient noise on speaker intelligibility for words and phrases. *The Journal of the Acoustical Society of America* **29**, 1320–1323.
Dufour, S. & Nguyen, N. (2013). How much imitation is there in a shadowing task? *Frontiers in Psychology* **4**. https://doi.org/10.3389/fpsyg.2013.00346.
Egnor, S. E. R. & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences* **27**, 649–654.
Egnor, S. E. R., Wickelgren, J. G. & Hauser, M. D. (2007). Tracking silence: adjusting vocal production to avoid acoustic interference. *Journal of Comparative Physiology A* **193**, 477–483.
Eliades, S. J. & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* **453**, 1102–1106.
Eliades, S. J. & Wang, X. (2012). Neural correlates of the Lombard effect in primate auditory cortex. *The Journal of Neuroscience* **32**, 10737–10748.
Evans, B. G. & Iverson, P. (2007). Plasticity in vowel perception and production: a study of accent change in young adults. *Journal of the Acoustical Society of America* **121**, 3814–3826.
Ey, E., Rahn, C., Hammerschmidt, K. & Fischer, J. (2009). Wild female olive baboons adapt their grunt vocalizations to environmental conditions. *Ethology* **115**, 493–503.
Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience* **15**, 399–402.
Farabaugh, S. M., Linzenbold, A. & Dooling, R. J. (1994). Vocal plasticity in budgerigars (*Melopsittacus undulatus*): evidence for social factors in the learning of contact calls. *Journal of Comparative Psychology* **108**, 81–92.
Fischer, J. (2017). Primate vocal production and the riddle of language evolution. *Psychonomic Bulletin & Review* **24**, 72–78.
Fischer, J., Hammerschmidt, K. & Todt, D. (1998). Local variation in Barbary macaque shrill barks. *Animal Behaviour* **56**, 623–629.
Fischer, J., Wheeler, B. C. & Higham, J. P. (2015). Is there any evidence for vocal learning in chimpanzee food calls? *Current Biology* **25**, R1028–R1029.
Fitch, T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences* **4**, 258–267.
Foulkes, P., Docherty, G., Khattab, G. & Yaeger-Dror, M. (2011). Sound judgements: perception of indexical features in children's speech. In *A Reader in Sociophonetics* (eds D. Preston and N. Niedzielski), pp. 327–356. de Gruyter Mouton, Berlin and New York.
Fowler, C. & Galantucci, B. (2005). The relation of speech perception and speech production. In *The Handbook of Speech Perception* (Volume , eds D. B. Pisoni and R. E. Remez), pp. 632–652. Blackwell Publishing, Malden.
Gambi, C. & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology* **4**. https://doi.org/10.3389/fpsyg.2013.00340.
Garland, E. C., Gedamke, J., Rekdahl, M. L., Noad, M. J., Garrigue, C. & Gales, N. (2013). Humpback whale song on the Southern Ocean feeding grounds: implications for cultural transmission. *PLoS ONE* **8**, e79422.
Garland, E. C., Goldizen, A. W., Rekdahl, M. L., Constantine, R., Garrigue, C., Hauser, N. D., Poole, M. M., Robbins, J. & Noad, M. J. (2011). Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. *Current Biology* **21**, 687–691.
Garnier, M. & Henrich, N. (2014). Speaking in noise: how does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech & Language* **28**, 580–597.
Garnier, M., Henrich, N. & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language and Hearing Research* **53**, 588–608.
Giles, H. (1973). Accent mobility: a model and some data. *Anthropological Linguistics* **15**, 87–105.
Giles, H. & Baker, S. (2008). Communication accommodation theory. In *The International Encyclopedia of Communication* (ed. W. Donsbach). Blackwell Publishing, Blackwell Reference Online, Oxford. http://www.blackwellreference.com/subscriber/tocnode.html?id=g9781405131995_yr2015_chunk_g97814051319958_ss68-1.
Giles, H., Coupland, J. & Coupland, N. (1991). Accommodation theory: communication, context, and consequence. In *Contexts of Accommodation: Developments in Applied Sociolinguistics* (eds H. Giles, J. Coupland and N. Coupland), pp. 1–68. Cambridge University Press, Cambridge.
Giles, H. & Powesland, P. F. (1975). *Speech Style and Social Evaluation*. Academic Press, London, New York.
Giles, H., Taylor, D. M. & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: some Canadian data. *Language in Society* **2**, 177–192.
Glaser, E. & Bart, G. (2015). Dialektsyntax des Schweizerdeutschen. In *Regionale Variation des Deutschen: Projekte und Perspektiven* (eds R. Kehrein, A. Lameli and S. Rabanus), pp. 81–107. De Gruyter, Berlin and Boston.
Glazko, G. V. & Nei, M. (2003). Estimation of divergence times for major lineages of primate species. *Molecular Biology and Evolution* **20**, 424–434.
Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review* **105**, 251–279.

Goldinger, S. D. & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review* **4**, 716–722.
Gregory, S., Dagan, K. & Webster, S. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior* **21**, 23–43.
Hage, S. R., Gavrilov, N. & Nieder, A. (2013*a*). Cognitive control of distinct vocalizations in rhesus monkeys. *Journal of Cognitive Neuroscience* **25**, 1692–1701.
Hage, S. R., Jiang, T., Berquist, S. W., Feng, J. & Metzner, W. (2013*b*). Ambient noise induces independent shifts in call frequency and amplitude within the Lombard effect in echolocating bats. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 4063–4068.
Hage, S. R., Jürgens, U. & Ehret, G. (2006). Audio–vocal interaction in the pontine brainstem during self-initiated vocalization in the squirrel monkey. *European Journal of Neuroscience* **23**, 3297–3308.
Hage, S. R. & Nieder, A. (2013). Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nature Communications* **4**, 2409.
Hage, S. R. & Nieder, A. (2016). Dual neural network model for the evolution of speech and language. *Trends in Neurosciences* **39**, 813–829.
Hammerschmidt, K. & Fischer, J. (2008). Constraints in primate vocal production. In *The Evolution of Communicative Creativity: From Fixed Signals to Contextual Flexibility* (eds. D. K. Oller and U. Griebel), pp. 93–119. MIT Press, Cambridge.
Haun, D. & Over, H. (2015). Like me: a homophily-based account of human culture. In *Epistemological Dimensions of Evolutionary Psychology* (ed. T. Breyer), pp. 117–130. Springer, New York.
Healey, P. G. T., Purver, M. & Howes, C. (2014). Divergence in dialogue. *PLoS ONE* **9**, e98598.
Heblich, S., Lameli, A. & Riener, G. (2015). The effect of perceived regional accents on individual economic behavior: a lab experiment on linguistic performance, cognitive ratings and economic decisions. *PLoS ONE* **10**, e0113475.
Holt, D. E. & Johnston, C. E. (2014). Evidence of the Lombard effect in fishes. *Behavioral Ecology* **25**. https://doi.org/10.1093/beheco/aru028.
Hopkins, W. D., Taglialatela, J. & Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour* **73**, 281–286.
Hotchkin, C. & Parks, S. (2013). The Lombard effect and other noise-induced vocal modifications: insight from mammalian communication systems. *Biological Reviews* **88**, 809–824.
Hotchkin, C. F., Parks, S. E. & Weiss, D. J. (2015). Noise-induced frequency modifications of tamarin vocalizations: implications for noise compensation in nonhuman primates. *PLoS ONE* **10**, e0130211.
Janik, V. & Slater, P. (2000). The different roles of social learning in vocal communication. *Animal Behaviour* **60**, 1–11.
Junqua, J.-C. (1996). The influence of acoustics on speech production: a noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication* **20**, 13–22.
Kim, M., Horton, W. & Bradlow, A. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Journal of Laboratory Phonology* **2**, 125–156.
King, S. L. & McGregor, P. K. (2016). Vocal matching: the what, the why and the how. *Biology Letters* **12**, 20160666.
Knörnschild, M. (2014). Vocal production learning in bats. *Current Opinion in Neurobiology* **28**, 80–85.
Koski, S. E. & Burkart, J. M. (2015). Common marmosets show social plasticity and group-level similarity in personality. *Scientific Reports* **5**. https://doi.org/10.1038/srep08878.
Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience* **5**, 831–843.
Kumar, V., Croxson, P. L. & Simonyan, K. (2016). Structural organization of the laryngeal motor cortical network and its implication for evolution of speech production. *Journal of Neuroscience* **36**, 4170–4181.
Lameira, A. R., Hardus, M. E., Mielke, A., Wich, S. A. & Shumaker, R. W. (2016). Vocal fold control beyond the species-specific repertoire in an orang-utan. *Scientific Reports* **6**, 30315.
Lane, H. & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech, Language and Hearing Research* **14**, 677–709.
Langergraber, K. E., Prüfer, K., Rowney, C., Boesch, C., Crockford, C., Fawcett, K., Inoue, E., Inoue-Muruyama, M., Mitani, J., Muller, M., Robbins, M., Schubert, G., Stoinski, T. S., Viola, B., Watts, D., Wittig, R., Wrangham, R. W., Zuberbühler, K., Pääbo, S. & Vigilant, L. (2012). Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 15716–15721.
Laporte, M. N. C. & Zuberbühler, K. (2010). Vocal greeting behaviour in wild chimpanzee females. *Animal Behaviour* **80**, 467–473.
Lemasson, A. & Hausberger, M. (2004). Patterns of vocal sharing and social dynamics in a captive group of Campbell's monkeys (*Cercopithecus campbelli campbelli*). *Journal of Comparative Psychology* **118**, 347–359.
Lemasson, A., Jubin, R., Masataka, N. & Arlet, M. (2016). Copying hierarchical leaders' voices? Acoustic plasticity in female Japanese macaques. *Scientific Reports* **6**. https://doi.org/10.1038/srep21289.
Lemasson, A., Ouattara, K., Petit, E. J. & Zuberbühler, K. (2011). Social learning of vocal structure in a nonhuman primate? *BMC Evolutionary Biology* **11**, 1–7.
Lemasson, A., Ouattara, K. & Zuberbühler, K. (2013). Exploring the gaps between primate calls and human language. In *The Evolutionary Emergence of Language: Evidence and Inference*, (eds. R. Brotha and M. Everaert) pp. 181–203. Oxford University Press, Oxford.
Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* **21**, 1–36.
Liebal, K., Waller, B. M., Burrows, A. M. & Slocombe, K. E. (2013). *Primate Communication: A Multimodal Approach*. Cambridge University Press, Cambridge.
Loh, K. K., Petrides, M., Hopkins, W. D., Procyk, E. & Amiez, C. (2016). Cognitive control of vocalizations in the primate ventrolateral-dorsomedial frontal (VLF-DMF) brain network. *Neuroscience & Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2016.12.001.
Lu, Y. & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America* **124**, 3261–3275.
MacLeod, B. (2012). *The effect of perceptual salience on phonetic accommodation in cross-dialectal conversation in Spanish*. PhD Thesis: University of Toronto, Toronto.
Massen, J. J. M. & Koski, S. E. (2014). Chimps of a feather sit together: chimpanzee friendships are based on homophily in personality. *Evolution and Human Behavior* **35**, 1–8.
Miller, C. T., Thomas, A. W., Nummela, S. U. & de la Mothe, L. A. (2015). Responses of primate frontal cortex neurons during natural vocal communication. *Journal of Neurophysiology* **114**. https://doi.org/10.1152/jn.01003.2014.
Mitani, J. C. & Gros-Louis, J. (1998). Chorusing and call convergence in chimpanzees: tests of three hypotheses. *Behaviour* **135**, 1041–1064.
Molesti, S. & Majolo, B. (2015). Cooperation in wild barbary macaques: factors affecting free partner choice. *Animal Cognition* **19**, 133–146.
Morrison, R. & Reiss, D. (2013). Whisper-like behavior in a non-human primate. *Zoo Biology* **32**, 626–631.
Morton, F. B., Weiss, A., Buchanan-Smith, H. M. & Lee, P. C. (2015). Capuchin monkeys with similar personalities have higher-quality relationships independent of age, sex, kinship and rank. *Animal Behaviour* **105**, 163–171.
Mundinger, P. C. (1970). Vocal imitation and individual recognition of finch calls. *Science* **168**, 480–482.
Mundinger, P. C. (1979). Call learning in the Carduelinae: ethological and systematic considerations. *Systematic Biology* **28**, 270–283.
Namy, L., Nygaard, L. & Sauerteig, D. (2002). Gender differences in vocal accommodation: the role of perception. *Journal of Language & Social Psychology* **21**, 422–432.
Nguyen, N. & Delvaux, V. (2015). Role of imitation in the emergence of phonological systems. *Journal of Phonetics* **53**, 46–54.
Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics* **39**, 132–142.
Nonaka, S., Takahashi, R., Enomoto, K., Katada, A. & Unno, T. (1997). Lombard reflex during PAG-induced vocalization in decerebrate cats. *Neuroscience Research* **29**, 283–289.
Nowicki, S. (1989). Vocal plasticity in captive black-capped chickadees: the acoustic basis and rate of call convergence. *Animal Behaviour* **37** Pt. 1, 64–73.
Owren, M. J., Amoss, R. T. & Rendall, D. (2011). Two organizing principles of vocal production: implications for nonhuman and human primates. *American Journal of Primatology* **73**, 530–544.
Paas, F., Renkl, A. & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist* **38**, 1–4.
Pardo, J. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* **119**, 2382–2392.
Pardo, J. (2012). Reflections on phonetic convergence: speech perception does not mirror speech production. *Language and Linguistics Compass* **6**, 753–767.
Pardo, J., Gibbons, R., Suppes, A. & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics* **40**, 190–197.
Pardo, J., Jay, I. C., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C. & Krauss, R. M. (2013*a*). Influence of role-switching on phonetic convergence in conversation. *Discourse Processes* **50**, 276–300.
Pardo, J., Jordan, K., Mallari, R., Scanlon, C. & Lewandowski, E. (2013*b*). Phonetic convergence in shadowed speech: the relation between acoustic and perceptual measures. *Journal of Memory and Language* **69**, 183–195.
Parks, S. E., Johnson, M., Nowacek, D. & Tyack, P. L. (2010). Individual right whales call louder in increased environmental noise. *Biology Letters* **7**. https://doi.org/10.1098/rsbl.2010.0451.
Patel, R. & Schell, K. W. (2008). The influence of linguistic content on the Lombard effect. *Journal of Speech, Language, and Hearing Research* **51**, 209–220.
Paukner, A., Suomi, S. J., Visalberghi, E. & Ferrari, P. F. (2009). Capuchin monkeys display affiliation toward humans who imitate them. *Science* **325**, 880–883.

Pelegrín-García, D., Smits, B., Brunskog, J. & Jeong, C.-H. (2011). Vocal effort with changing talker-to-listener distance in different acoustic environments. *The Journal of the Acoustical Society of America* **129**, 1981–1990.
Pharao, N., Maegaard, M., Møller, J. S. & Kristiansen, T. (2014). Indexical meanings of [s+] among Copenhagen youth: social perception of a phonetic variant in different prosodic contexts. *Language in Society* **43**, 1–31.
Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R. & Kearney, J. K. (1989). Inhibiting the Lombard effect. *The Journal of the Acoustical Society of America* **85**, 894–900.
Pickering, M. J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral & Brain Sciences* **27**, 169–190.
Pickering, M. J. & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* **36**, 329–347.
Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S. & Watwood, S. (2005). Animal behaviour: elephants are capable of vocal learning. *Nature* **434**, 455–456.
Prather, J. F., Peters, S., Nowicki, S. & Mooney, R. (2008). Precise auditory–vocal mirroring in neurons for learned vocal communication. *Nature* **451**, 305–310.
Ralls, K., Fiorelli, P. & Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. *Canadian Journal of Zoology* **63**, 1050–1056.
Reubold, U. & Harrington, J. (2015). Dissociating the effects of age from phonetic change. In *Language Development: The Lifespan Perspective* (eds A. Gerstenberg and A. Voeste), pp. 9–38. John Benjamins Publishing Company, Amsterdam and Philadelphia.
Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., Mathew, S., Newton, E. K., Naar, N., Newson, L., Ross, C., Smaldino, P. E., Waring, T. M. & Zefferman, M. (2016). Cultural group selection plays an essential role in explaining human cooperation: a sketch of the evidence. *Behavioral and Brain Sciences* **39**, e30.
Ridgway, S., Carder, D., Jeffries, M. & Todd, M. (2012). Spontaneous human speech mimicry by a cetacean. *Current Biology* **22**, R860–R861.
Roy, S., Miller, C., Gottsch, D. & Wang, X. (2011). Vocal control by the common marmoset in the presence of interfering noise. *Journal of Experimental Biology* **214**, 3619–3629.
Rukstalis, M., Fite, J. E. & French, J. A. (2003). Social change affects vocal structure in a callitrichid primate (*Callithrix kuhlii*). *Ethology* **109**, 327–340.
Sancier, M. & Fowler, C. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* **25**, 421–436.
Sanvito, S., Galimberti, F. & Miller, E. H. (2007). Observational evidences of vocal learning in southern elephant seals: a longitudinal study. *Ethology* **113**, 137–146.
Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K. & Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS ONE* **8**, e76674.
Schweitzer, A. & Lewandowski, N. (2013). Convergence of articulation rate in spontaneous speech. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, pp. 525–529. Lyon. 25–29 August 2013.
Schweitzer, A. & Lewandowski, N. (2014). Social factors in convergence of F1 and F2 in spontaneous speech. In *Proceedings of the 10th International Seminar on Speech Production*, p. Cologne, Germany. 5–8 May 2014.
Sewall, K. B., Young, A. M. & Wright, T. F. (2016). Social calls provide novel insights into the evolution of vocal learning. *Animal Behaviour* **120**, 163–172.
Seyfarth, R. M. & Cheney, D. L. (1997). Some general features of vocal development in nonhuman primates. In *Social Influences on Vocal Development* (eds C. T. Snowdon and M. Hausberger), pp. 249–273. Cambridge University Press, Cambridge.
Shih, C. & Lu, H.-Y. D. (2015). Effects of talker-to-listener distance on tone. *Journal of Phonetics* **51**, 6–35.
Shockley, K., Sabadini, L. & Fowler, C. (2004). Imitation in shadowing words. *Perception & Psychophysics* **66**, 422–429.
Siegel, J. (2010). *Second Dialect Acquisition*. Cambidge University Press, Cambridge.
Silk, J. B. (2014). Evolutionary perspectives on the links between close social bonds, health, and fitness. In *Sociality, Hierarchy, Health: Comparative Biodemography* (eds M. Weinstein and M. A. Lane), pp. 121–143. National Academies Press, Washington.
Skipper, J. I., Devlin, J. T. & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: review of the role of the motor system in speech perception. *Brain and Language* **164**, 77–105.
Smolker, R. & Pepper, J. W. (1999). Whistle convergence among allied male bottlenose dolphins (Delphinidae, *Tursiops* sp.). *Ethology* **105**, 595–617.
Snowdon, C. T. (2001). Social processes in communication and cognition in callitrichid monkeys: a review. *Animal Cognition* **4**, 247–257.
Snowdon, C. T. (2013). Language parallels in new world primates. In *Animal Models of Speech and Language Disorders* (ed. S. A. Helekar), pp. 241–261. Springer, New York.
Snowdon, C. T. (2017). Learning from monkey "talk". *Science* **355**, 1120–1122.
Snowdon, C. T. & Elowson, M. A. (1999). Pygmy marmoset modify their call structure when paired. *Ethology* **105**, 782–897.
Sonderegger, M., Bane, M. & Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language* **93**, 598–640.
Stoeger, A. S., Mietchen, D., Oh, S., de Silva, S., Herbst, C. T., Kwon, S. & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology* **22**, 2144–2148.
Stowe, L. M. & Golob, E. J. (2013). Evidence that the Lombard effect is frequency-specific in humans. *The Journal of the Acoustical Society of America* **134**, 640–647.
Sugiura, H. (1998). Matching of acoustic features during the vocal exchanges of coo calls in Japanese macaques. *Animal Behaviour* **55**, 673–687.
Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I. & Stokes, M. A. (1988). Effects of noise on speech production: acoustic and perceptual analyses. *The Journal of the Acoustical Society of America* **84**, 917–928.
Taglialatela, J. P., Russell, J. L., Schaeffer, J. A. & Hopkins, W. D. (2008). Communicative signaling activates 'Broca's' homolog in chimpanzees. *Current Biology* **18**, 343–348.
Taglialatela, J. P., Russell, J. L., Schaeffer, J. A. & Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. *PLoS ONE* **6**, e18852.
Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie* **20**, 410–433.
Tomasello, M. (2008). *Origins of Human Communication*. MIT Press, Cambridge.
Trudgill, P. (1986). *Dialects in Contact*. Blackwell Publishing, Oxford.
Trudgill, P. (2002). *Sociolinguistic Variation and Change*. Georgetown University Press, Washington.
Trudgill, P. (2008). Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society* **37**, 241–254.
Tyack, P. L. (2008). Convergence of calls as animals form social bonds, active compensation for noisy communication channels, and the evolution of vocal learning in mammals. *Journal of Comparative Psychology (Washington, D.C.: 1983)* **122**, 319–331.
Tyack, P. L. (2016). Vocal learning and auditory-vocal feedback. In *Vertebrate Sound Production and Acoustic Communication* (eds R. A. Suthers, W. T. Fitch, R. R. Fay and A. N. Popper), pp. 261–295. Springer, Cham.
Walker, A. & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology* **6**. https://doi.org/10.3389/fpsyg.2015.00546.
Walker, A., García, C., Cortés, Y. & Campbell-Kibler, K. (2014). Comparing social meanings across listener and speaker groups: the indexical field of Spanish /s/. *Language Variation and Change* **26**, 169–189.
Watkins, K. E., Strafella, A. P. & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* **41**, 989–994.
Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V. & Slocombe, K. E. (2015*a*). Reply to Fischer et al. *Current Biology* **25**, R1030–R1031.
Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V. & Slocombe, K. E. (2015*b*). Vocal learning in the functionally referential food grunts of chimpanzees. *Current Biology* **25**, 495–499.
Werlen, I. & Schlegel, D. (2006). *Zwischen 'Grüessech' und 'Tagwoll'. Das Sprachverhalten und die Lebenssituation der Oberwalliser und Oberwalliserinnen in Bern*. Institut für Sprachwissenschaft, Universität Bern, Bern.
Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E. & Shumaker, R. W. (2008). A case of spontaneous acquisition of a human sound by an orangutan. *Primates* **50**, 56–64.
Wilson, M. L., Hauser, M. D. & Wrangham, R. W. (2007). Chimpanzees (*Pan troglodytes*) modify grouping and vocal behaviour in response to location-specific risk. *Behaviour* **144**, 1621–1653.
Yu, A. C. L., Abrego-Collier, C. & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: subjective attitude, personality and "autistic" traits. *PLoS ONE* **8**, e74746.
Zuberbühler, K. (2011). Cooperative breeding and the evolution of vocal flexibility. In *The Oxford Handbook of Language Evolution* (eds M. Tallerman and K. R. Gibson), pp. 71–81. Oxford University Press, Oxford and New York.
Zuberbühler, K. (2015). Linguistic capacity of non-human animals. *Wiley Interdisciplinary Reviews: Cognitive Science* **6**, 313–321.
Zürcher, Y. & Burkart, J. M. (2017). Evidence for dialects in three captive populations of common marmosets (*Callithrix jacchus*). *International Journal of Primatology* **38**, 780–793.

(*Received 17 July 2017; revised 26 September 2017; accepted 2 October 2017*)